MODELISATION ET PREVISION DES DEBITS NATURELS JOURNALIERS DU B.V.I. DE LA SANAGA A LA STATION DE CONTRÔLE DE SONGMBENGUE

Par :
DKENGNE SIELENOU Pascal
Maître ès Sciences

Dirigé par :
Dr. Eugène-Patrice NDONG NGUEMA
Chargé de Cours à l'Ecole Nationale Supérieure Polytechnique de Yaoundé

et de

M. ATANGANA Martin

Analyste Junior Chargé du Risque Hydrologique à AES-SONEL de Douala

Octobre 2006

Dédicaces

Je dédie ce mémoire à tous les membres de la famille SIELENOU et de la famille DJAMBONG pour vos sacrifices, vos encouragements et pour votre soutien permanent. Recevez ce travail en guise de reconnaissance pour votre amour et labeur que vous n'avez cessé de m'apporter.

Remerciements

Eu égard aux nombreuses difficultés relatives à ce sujet, ce mémoire serait encore un rêve aujourd'hui, si je n'avais pas bénéficié de l'assistance morale et scientifique de nombreuses personnalités et amis, à qui je tiens à exprimer la marque de ma profonde gratitude.

Je tiens à témoigner ma profonde gratitude à M. ATANGANA Martin, analyste junior chargé du risque hydrologique à AES-SONEL, au Dr. Eugène-Patrice NDONG NGUEMA qui, malgré leurs nombreuses occupations, étaient toujours disponibles à suivre ce travail.

Ma reconnaissance va naturellement au professeur Henri GWET et à M. Isidore PRISO KOTTO, qui ont permis que je fasse mon stage à AES-SONEL.

Ma reconnaissance et mes sincères remerciements vont à la collaboration AMMSI (African Mathematics Millenium Science Initiative) pour la main forte qu'elle m'a prêtée pour la rédaction de ce mémoire à travers le professeur David BEKOLLE.

Je tiens également à exprimer ma reconnaissance et mes sincères remerciements à Mme KENNE Emilie, Mme MAPA Sidonie, Mlle Alvine, TENKEU Aldoph, TAKAM Patrice, MAHIANE Guy, WHEGANG Solange, KOUAMO Olaf, TOUSSILE Wilson, DJIMEFO Georges Parfait, TCHIDA Emmanuel, FOTSING Réné, tous mes camarades de promotion et enfin au Seigneur tout puissant à qui je dois tout.

Lexique et abréviations

- **Hydrologie :** science qui étudie les propriétés de l'eau et sa circulation naturelle à la surface de la terre ;
- **Module d'un cours d'eau :** c'est la valeur du débit moyen annuel de ce cours d'eau calculée sur une longue chronique ;
- **Période de basses eaux, période d'étiage ou saison sèche :** c'est la période de l'année où le débit naturel d'un fleuve est inférieur à son module ;
- **Période de hautes eaux, de crue ou saison humide :** c'est la période de l'année où le débit naturel du fleuve est supérieur à son module ;
- **Année moyenne :** année au cours de laquelle une variable hydrologique observée a une valeur voisine de la moyenne de cette variable calculée sur une longue chronique ;
- **Année hydrologique :** période continue de 12 mois alternant une période de crue et une période d'étiage ;
- **Bassin versant d'un cours d'eau :** il correspond à la surface au sol où toute précipitation tombée va alimenter ce cours d'eau et l'ensemble de ses affluents. Par extension, toute précipitation tombant à l'extérieur du bassin versant alimentera d'autres cours d'eau ;
- **Bilan hydrique ou hydrologique :** bilan d'eau fondé sur le principe que, pendant un certain intervalle de temps, le total des apports à un bassin versant doit être égal au total des sorties plus la variation, positive ou négative, du volume d'eau stocké dans le dit bassin ;
- **Bassin versant d'un barrage réservoir :** il correspond à la surface au sol où toute précipitation tombée va alimenter ce réservoir. Par extension, toute précipitation tombant à l'extérieur du bassin versant alimentera d'autres cours d'eau ne se dirigeant pas vers ce réservoir ;
- **Bassin versant intermédiaire (B.V.I.) en amont d'un point situé en aval des barrages réservoirs :** c'est la superficie totale de la zone située entre ce point et les barrages réservoirs. Le débit naturel du fleuve en ce point n'est plus tout à fait un débit naturel mais un **débit naturel reconstitué** ;

- **Régime hydrologique d'un cours d'eau :** c'est l'ensemble des caractéristiques hydrologiques de ce cours d'eau et son mode de variation ;
- **Le bassin versant intermédiaire de la Sanaga en amont de la station de contrôle de Songmbengué :** c'est la superficie totale de la zone située entre cette station et l'ensemble des trois barrages de retenue situés en amont ;
- **Débit d'un cours d'eau :** représente le volume total d'eau qui s'écoule à travers une section transversale de ce cours d'eau pendant l'unité de temps. Il s'exprime en m^3/s ;
- **Jaugeage :** ensemble des opérations de calibrage d'un cours d'eau ayant pour but de déterminer une relation entre la hauteur et le débit de ce cours d'eau ;
- **Période de régularisation :** c'est la période pendant laquelle le débit du fleuve est soutenu par les apports provenant des lâchers effectués à partir des barrages réservoirs ;
- **Débit régularisé :** c'est le débit mesuré du fleuve pendant la période de régularisation ; c'est dont le débit naturel amélioré par des quantités d'eau émanant des réservoirs ;
- **Débit régularisé visé :** en fonction de la demande en énergie, de la situation du parc de production (puissance installée aux usines de production et aux centrales thermiques), le gestionnaire des aménagements fixe un débit régularisé en début de saison sèche qu'il s'estime pouvoir fournir ;
- **Débit mesuré :** c'est le débit tel qu'on l'observe réellement en un point donné d'un cours d'eau ;
- **Eclusée :** c'est la quantité d'eau qui coule entre le moment où l'on ouvre les portes d'une écluse et le moment où on les referme.
- **Débit naturel :** le débit d'un cours d'eau est dit naturel lorsqu'il correspond au régime d'écoulement du fleuve non perturbé par la présence de retenues artificielles telles que les barrages réservoirs. Lorsque le fleuve n'est pas équipé, le débit naturel correspond au débit mesuré. En revanche, lorsqu'il s'agit d'un fleuve équipé de barrages réservoirs stockant et lâchant des stocks considérables d'eau, le terme « débit naturel » n'a plus de signification physique ;
- **Débit naturel reconstitué d'un fleuve :** c'est le débit qu'aurait parcouru le fleuve si celui-ci n'avait pas été équipé. Ce débit n'est pas mesurable. Son estimation tient compte des données suivantes :
    a) Etat des stocks des réservoirs au début de la période de calculs ;
    b) Volume et durée des lâchers d'eau ;
    c) Volume et durée des rétentions d'eau dans les réservoirs ;

d) Pertes dues à l'exploitation (temps de manœuvre des organes, évaporation, fuite, etc.) ;
e) Apports dus aux pluies sur le bassin versant intermédiaire, c'est-à-dire entre les barrages réservoirs et le point où l'on cherche à calculer le débit reconstitué ;
f) Comportement et évolution des volumes d'eau lâchés le long du parcours.

Ainsi, le débit naturel reconstitué est le résultat d'un calcul complexe faisant appel à un très grand nombre de mesures différentes.

- **Turbine :** c'est un dispositif rotatif destiné à utiliser la force vive de l'eau et à transmettre le mouvement à un alternateur pour qu'il soit transformé en énergie électrique ;
- **Usine de production d'hydroélectricité :** c'est une usine qui permet de transformer l'énergie de chute d'un cours d'eau, c'est-à-dire la variation brusque d'altitude d'une masse d'eau, en énergie mécanique à l'aide d'une ou de plusieurs turbines, puis en énergie électrique à travers le rotor et le stator d'un générateur ;
- **Puissance installée :** c'est la somme totale des puissances maximales des turbines de la centrale. L'unité est le Méga Watt (MW) ;

Quelques abréviations

AES : All Energy Service ;

SONEL : Société Nationale d'Electricité ;

B.V.I. : Bassin Versant Intermédiaire ;

AR : AutoRegressive (autorégressif) ;

MA : Moving Average (moyenne mobile) ;

ARMA : AutoRegressive Moving Average (autorégressif moyenne mobile) ;

ARIMA : AutoRegressive Integrated Moving Average (autorégressif moyenne mobile intégré) ;

SARIMA : Seasonal AutoRegressive Integrated Moving Average (autorégressif moyenne mobile intégré saisonnier) ;

er : erreur quadratique relative moyenne ;

em : erreur quadratique moyenne.

Résumé

L'augmentation de la population Camerounaise, ainsi que la création de nouvelles activités imposent à AES-SONEL, seul fournisseur d'énergie électrique du pays, de prendre des dispositions pour assurer la qualité et la continuité de son service et ceci, en limitant au maximum ses dépenses. Elle doit, par conséquent, optimiser la gestion de ses ressources hydriques.

Le travail que nous avons effectué à la Direction des Etudes et de la Planification Stratégique de la société AES-SONEL est basé sur une analyse statistique des débits naturels journaliers du bassin versant intermédiaire de la Sanaga à la station de contrôle de Songmbengué.

En première analyse, nous avons standardisé cette série chronologique, puis nous avons modélisé la série résultante par le modèle $ARIMA(11,1,1)$. En seconde analyse, nous avons effectué une modélisation de ces débits par un modèle de type multiplicatif, et ceci après une désaisonnalisation de cette série à l'aide d'une moyenne mobile. En vue de comparer les résultats de ces deux modélisations, les erreurs quadratiques relatives moyennes (er) et les erreurs quadratiques moyennes (em) de la prévision a posteriori de l'ajustement ont été évaluées. Ces modèles paraissent équivalents en termes de ces deux erreurs. En troisième analyse, nous avons modélisé les débits moyens mensuels correspondants à ces débits naturels journaliers par le modèle $SARIMA(1,0,0)(1,1,1)_{12}$. Enfin, en quatrième analyse, nous avons établi un modèle de simulation basé sur un processus markovien du premier ordre.

Ce travail va ainsi permettre à la société AES-SONEL, à partir d'un programme de prévision annuelle des débits naturels journaliers, réalisé à base du modèle de prévision retenu, de :

- maîtriser le coût de la dépense en volume de combustibles nécessaire pour la production de l'énergie électrique à partir des centrales thermiques ;
- planifier l'exploitation courante du système hydroélectrique en vue de répondre aux besoins de la production et de respecter certaines contraintes (inondation, environnement, etc.).

Abstract

The increase of the Cameroonian population, as well as the creation of new activities impose to AES-SONEL, only supplier of electric energy of the country to take arrangements to assure the quality and the continuity of its service and this while limiting to the maximum its expenses. It must optimize the management of its water resources therefore.

The work that we did in the Direction of the Studies and the Scheduling of the AES-SONEL society is based on a statistical analysis of the daily natural flow of the basin pouring intermediate of the Sanaga in the station of control of SONGMBENGUE.

In first analysis, we standardized this time series, then we have modelled the resulting time series by a process $ARIMA(11,1,1)$. In a second analysis, we did a modelling of these flows by a multiplicative composition diagram and this after elimination of the seasonal component of this time series by the moving average method. Quadratic relative averages errors (er) and the quadratic averages errors (em) of the adjustment and the a posteriori forecasting have been estimated. The first model is kept as more reliable, in terms of these two errors when compared to the second model. In a third analysis, we have modelled the monthly average flows corresponding to these daily natural flows, by the model $SARIMA(1,0,0)(1,1,1)_{12}$. In a fourth analysis, we established a model of simulation based on a markovian process of the first order.

This work is going to allow AES-SONEL thus, through a program of yearly forecasting of the daily natural flows, on the basis of the retained forecasting model:

- to master the cost of expense in necessary fuel volume for the production of electricity from the thermal power stations;
- to plan the daily exploitation of the hydroelectric system to meet answering the needs of the production and to respect some constraints (flooding, environment, etc.).

Table des matières

Dédicaces ---------- ii
Remerciements ---------- **iii**
Lexique et abréviations----------**iv**
Résumé----------**vii**
Abstract----------**viii**
Table des figures----------**11**
Liste des tableaux----------**13**
Introduction générale ----------**14**
Résumé exécutif----------**15**

Chapitre 1 Contexte et problématique----------**20**
1.1 Présentations des activités de AES-SONEL ----------**20**
1.1.1 Activité de production de l'énergie électrique----------21
1.1.2 Activité de transport de l'énergie électrique----------21
1.2 Présentation du bassin versant de la Sanaga-- ----------**21**
1.3 Aménagements hydroélectriques de la Sanaga----------22
1.3.1 Les ouvrages de production d'énergie électrique ----------22
1.3.2 Les ouvrages de régularisation du débit de la Sanaga ----------22
1.3.3 Les stations hydrologiques du B.V.I. de la Sanaga----------23
1.4 Régularisation de la Sanaga ----------**25**
1.5 Problématique Générale----------**28**

Chapitre 2 Identification et description des données ----------**29**
2.1 Identification et obtention des données----------**29**
2.2 Description des données----------30

Chapitre 3 Outils statistiques----------33

3.1 Modèles déterministes ----------33

3.2 Un algorithme de modélisation d'une chronique----------35

3.3 Modèles stochastiques ----------38

3.4 Méthodologie: modèle $ARIMA(p,d,q)$ ----------41

3.4.1 Identification à priori et estimation----------41

3.4.2 Vérification à posteriori et choix du modèle----------42

3.5 Un modèle de simulation: modèle markoviens du premier ordre----------43

3.6 Evaluation d'un modèle----------44

3.6.1 Erreurs d'ajustement et de prévision----------44

3.6.2 Une méthodologie d'évaluation à posteriori d'un modèle----------45

Chapitre 4 Applications et résultats ----------47

4.1 Modélisation et prévisions des débits naturels journaliers par un processus ARIMA----------47

4.2 Modélisation et prévision des débits naturels journaliers par un modèle multiplicatif----------59

4.3 Modélisation des débits moyens mensuels par un modèle de type SARIMA-----67

Conclusion générale ----------74

Annexe----------75

Bibliographie----------101

Table des figures

1.1 Système de production hydroélectrique de la Sanaga----------24
1.2 Stations hydrologiques de la Sanaga----------27
2.1 Evolution des débits naturels journaliers----------30
2.2 Evolution journalière des débits naturels des années 1988 à 1994----------31
2.3 Evolution journalière des débits naturels des années 1994 à 2000----------31
2.4 Evolution journalière des débits naturels des années 2000 à 2005----------32
2.5, 2.6, 2.7 Débits du moi de Décembre par année----------76
2.8, 2.9, 2.10 Débits du moi de Janvier par année----------77
2.11, 2.12, 2.13 Débits du moi de Février par année----------78
2.14, 2.15, 2.16 Débits du moi de Mars par année----------79
2.17, 2.18, 2.19 Débits du moi d'Avril par année----------80
2.20, 2.21, 2.22 Débits du moi de Mai par année----------81
2.23, 2.24, 2.25 Débits du moi de Juin par année----------82
2.26, 2.27, 2.28 Débits du moi de Juillet par année----------83
2.29, 2.30, 2.31 Débits du moi d'Août par année----------84
2.32, 2.33, 2.34 Débits du moi de Septembre par année----------85
2.35, 2.36, 2.37 Débits du moi d'Octobre par année----------86
2.38, 2.39, 2.40 Débits du moi de Novembre par année----------87
4.1 Evolution des débits standardidés----------48
4.2 Corrélogramme des débits standardisés----------48
4.3 ACF des débits standardisés et différenciés----------49
4.4 Evolution des débits standardisés et différenciés----------50
4.5 ACF des débits standardisés et différenciés 2 fois----------50
4.6 PACF des débits standardisés et différenciés----------51
4.7 Diagnostic des résidus ε_t du modèle $ARIMA(11,1,1)$----------55
4.8 Histogramme des résidus standardisés----------56
4.9 Ajustement de la série des débits $X_{\nu,\tau}$ par le modèle final----------57

4.10 Courbe des débits réels et des prévisions----------58
4.11 Série des moyennes mobiles d'ordre 365----------60
4.12 Graphe de la série corrigée des variations saisonnières----------60
4.13 Graphe des résidus----------61
4.14 ACF des résidus estimés----------62
4.15 Corrélogrammes du processus w_t ----------62
4.16 Diagnostic des résidus η_t ----------64
4.17 Ajustement des débits y_t ----------65
4.18 Courbes des débits réels et des prévisions----------66
4.19 Courbe des débits moyens mensuels----------67
4.20 Autocorrélogrammes du processus Y_t ----------68
4.21 Autocorrélogrammes du processus Z_t ----------68
4.22 Diagnostic des résidus ε_t ----------70
4.23 Ajustement des débits moyens mensuels----------71
4.24, 4.25 Courbe des débits journaliers simulés----------73

Liste des tableaux

2.1 Débits naturels journaliers du B.V .I. de la Sanaga à Songmbengué----------102

2.2 Série des débits moyens mensuels du B.V.I. de la Sanaga à la station de contrôle de Songmbengué----------75

4.1 Paramètres estimés du modèle $ARIMA(14,1,1)$ ----------53

4.2 Paramètres estimés du modèle $ARIMA(13,1,1)$ ----------53

4.3 Paramètres estimés du modèle $ARIMA(12,1,1)$ ----------54

4.4 Paramètres estimés du modèle $ARIMA(11,1,1)$ ----------54

4.5 Erreurs quadratiques de l'ajustement du modèle final aux débits $X_{\nu,\tau}$ ----------57

4.6 Erreurs quadratiques des ajustements et des prévisions à posteriori----------58

4.7 Paramètres estimés du modèle $ARIMA(16,1,1)$ ----------63

4.8 Erreurs quadratiques de l'ajustement du modèle final aux débits y_t ----------65

4.9 Erreurs quadratiques des ajustements et des prévisions à posteriori----------66

4.10 Paramètres estimés du modèle $SARIMA(1,0,2)(1,1,1)_{12}$ ----------69

4.11 Paramètres estimés du modèle $SARIMA(1,0,1)(1,1,1)_{12}$ ----------70

4.13 Erreurs quadratiques de l'ajustement du modèle aux débits moyens mensuels Y_t ----------71

4.14 Coefficients saisonniers corrigés----------100

INTRODUCTION GENERALE

L'hydrologie est la science qui traite de la présence et de la distribution des eaux sur la terre, et leurs propriétés chimiques, biologiques et physiques, ainsi que leur interaction avec l'environnement physique. En tant que telle, elle est le fondement de la résolution des problèmes pratiques d'inondation, de sécheresse, d'érosion, de transport des sédiments, de la pollution des eaux et de la production de l'énergie.

L'énergie électrique est l'un des facteurs les plus importants de développement économique de la société moderne. L'énergie électrique d'origine hydraulique est une source d'énergie renouvelable. Les ouvrages de production hydroélectrique varient selon le site, le climat et la demande en énergie.

En ce qui concerne les ouvrages de production hydroélectrique de AES-SONEL, la connaissance à l'avance des débits naturels journaliers de la Sanaga juste à l'entrée des usines hydroélectrique est essentielle pour programmer l'exploitation des eaux stockées dans les barrages réservoirs situés en amont de ces usines. Mais, dans la pratique, fournir ce genre d'information est loin d'être aussi simple. L'objet de notre étude à la Direction de l'Etude et de la Planification de la société AES-SONEL est basé sur la modélisation et la prévision des débits naturels journaliers du Bassin Versant Intermédiaire de la Sanaga en amont de la station hydrologique de Songmbengué.

Notre étude s'articule autour de quatre chapitres :

- au premier chapitre, nous présentons le contexte et la problématique ;
- le deuxième chapitre quant à lui, est une présentation et une description des données ;
- le troisième chapitre repose sur un rappel de quelques notions statistiques indispensables à la réalisation de cette étude ;
- le quatrième chapitre est réservé à l'application des différentes méthodes présentées au troisième chapitre aux données et à la présentation des différents résultats obtenus.

Résumé exécutif

1. **SUJET**

Modélisation et prévision des débits naturels journaliers du Bassin Versant Intermédiaire (B.V.I.) de la Sanaga à la station de contrôle de Songmbengué.

2. PROBLEME

Au Cameroun, la Société AES-SONEL, produit de l'énergie électrique au moyen d'un système hydraulique et d'un système thermique.

La production d'énergie électrique d'origine hydraulique permettant d'alimenter le réseau Sud, c'est-à-dire les provinces du Centre, du Sud, du Littoral, de l'Ouest, du Nord-Ouest, et du Sud-Ouest, est faite dans les centrales hydroélectriques de Song-Loulou et d'Edéa. Le débit de l'eau alimentant ces centrales dépend directement des débits naturels du fleuve Sanaga à la station de contrôle de Songmbengué et des lâchers d'eau effectués au niveau des barrages réservoirs de Mbakaou, de Bamendjin et de Mapé. Le barrage de Mbakaou est situé à 7 jours (temps de propagation des éclusés) de la centrale de Song-Loulou, tandis que les barrages de Mapé et de Bamendjin sont situés à 5 jours de la centrale de Song-Loulou. Généralement, les vannes du barrages de Mbakaou sont maintenues ouvertes pendant une certaine période (en jours), de sorte que le débit de l'eau y sortant est constant sur toute la période. La question qui se pose donc ici est celle de savoir quelles sont les quantités d'eau à lâcher des barrages de Mapé et de Bamendjin, qui viendront s'ajouter 5 jours plus tard au débit naturel de la Sanaga à Songmbengué augmenté du débit constant de l'eau provenant du Barrage de Mbakaou, pour alimenter la centrale de Song-Loulou puis, celle d'Edéa afin de produire une quantité d'énergie électrique juste nécessaire pour satisfaire la demande tout en réduisant au maximum les pertes en eaux ? Mais, une planification des quantités d'eau à lâcher des barrages réservoirs ne pourra être parfaite que si l'on dispose des « bonnes » prévisions des débits naturels de la Sanaga à Songmbengué.

Par ailleurs, le coût de la dépense en volume de combustibles nécessaire pour la production d'énergie électrique à partir des centrales thermiques, a une proportion non négligeable dans le budget de AES-SONEL. Ce qui implique une planification annuelle de

cette dépense. Pour y parvenir, AES-SONEL a établi une procédure leur permettant d'estimer ce coût. Mais, parmi les données les plus importantes dont à besoin cette procédure pour fonctionner, figure une prévision sur un an des débits naturels journaliers à Songmbengué.

Il en résulte de tout ce qui précède qu'un programme permettant de faire une prévision annuelle des débits naturels journaliers de la Sanaga à Songmbengué est indispensable pour AES-SONEL.

3. DONNEES

En vue de déterminer le programme de prévision dont il est question dans cette étude, Nous allons d'abord chercher à modéliser (caractériser) la chronique des débits naturels reconstitués et journaliers de la Sanaga à Songmengué, calculés sur la période allant du 1er Juillet 1988 au 30 Juin 2005.

4. METHODOLOGIE

Nous avons utilisés dans notre analyse deux grandes méthodologies statistiques relatives aux séries temporelles pour essayer de modéliser cette chronique :

a- Pour la première démarche, nous avons d'abord standardiser cette chronique puis, nous avons appliqué à la chronique résultante une méthodologie proposée par BOX et JENKINS en vue de caractériser cette dernière chronique par un modèle $ARIMA(p,d,q)$.

b- Pour la seconde démarche, nous avons caractériser cette chronique des débits naturels journaliers à Songmbengué par un modèle multiplicatif, tout en lui appliquant une moyenne mobile centrée pour la désaisonnaliser.

5. RESULTATS

Il ressort de nos analyses que la chronique de nos débits naturels journaliers à Songmbengué peut être caractérisée par l'une des équations $(*)$ et $(**)$ suivantes :

$$(*) \qquad X_{\nu,\tau} = \sigma_\tau Z_{\nu,\tau} + \mu_\tau \ ,$$

$\nu = 1,2,...,17$; $\tau = 1,2,...,365$; où $X_{\nu,\tau}$ est le débit du jour numéro τ de l'année numéro ν ; μ_τ est la moyenne empirique des débits du jour numéro τ sur les 17 années d'observations ; σ_τ^2 est la variance empirique des débits du jour numéro τ sur les 17 années d'observations ; $Z_{\nu,\tau}$ est un processus qui suit le modèle $ARIMA(11,1,1)$ dont les paramètres sont stockés dans le tableau 4.7 du mémoire.

$$(**) \qquad y_t = (\alpha + \beta t) \times \left(\sum_{i=1}^{365} \gamma_i S_t^i \right) \times \varepsilon_t \ ,$$

avec $S_t^i = \begin{cases} 1, \text{ si } t \text{ correspond au jour i} \\ 0, \text{ sinon} \end{cases}$, où $\alpha = 1751$; $\beta = 0.0059$; les coefficients γ_i sont stockés dans le tableau 4.22 du mémoire; ε_t est un processus qui suit le modèle $ARIMA(16,1,1)$ dont les paramètres sont stockés dans le tableau 4.11 du mémoire.

Par ailleurs, la chronique des débits moyens mensuels correspondants aux débits naturels journaliers de la Sanaga à Songmbengué peut être caractérisée par le modèle $SARIMA(1,0,0)(1,1,1)_{12}$ dont les paramètres sont stockés dans le tableau 4.19 du mémoire.

6. CONCLUSION GENERALE

Nous avons soulevé au départ le problème de la détermination d'un programme permettant de faire des prévisions journalières des débits naturels du fleuve Sanaga à Songmbengué sur une période d'un an.

A cet effet, nous avons commencé par caractériser la chronique des débits naturels journaliers à Songmbengué par 2 modèles distincts définis par les équations $(*)$ et $(**)$. Ces 2 modèles sont quasiment équivalents en termes de leurs erreurs quadratiques d'ajustement et de prévision a posteriori. Notons que les erreurs quadratiques relatives moyennes d'adéquation de chacun de ces 2 modèles aux données ont une moyenne de 5%, tandis que les erreurs quadratiques relatives moyennes de prévisions des débits naturels journaliers d'une année à venir ont une moyenne de 23%.

Nous avons ensuite mis au point un programme R dont l'algorithme s'appuie sur le modèle défini par l'équation $(*)$. Ce programme pourra désormais permettre à AES-SONEL, de faire des prévisions des débits naturels journaliers à Songmbengué sur une période d'un an. Le mode d'emploi de ce programme de prédiction se trouve directement à la suite de ce résumé.

Nous ne saurons terminer ce résumé sans toute fois signaler que nous avons aussi mis au point dans cette étude un programme R dont l'algorithme s'appuie sur un processus de Markov du 1er ordre qui pourra permettre à AES-SONEL, d'effectuer des simulations des débits naturels journaliers de la Sanaga à Songmbengué.

MODE D'EMPLOI DU PROGRAMME DE PREVISION DES DEBITS NATURELS JOURNALIERS DU B.V.I. DE LA SANAGA A LA STATION DE CONTRÔLE DE SONGMBENGUE

1- Téléchargement et installation du logiciel R

Le logiciel R est une version non commerciale de S-plus. Télécharger ce logiciel sur le site Internet http://lib.stat.cmu.edu/R/CRAN/ ou sur le site http://cran.ch.r-project.org/, installer ensuite ce logiciel dans un ordinateur.

2- Création d'un répertoire de travail

a- Créer un dossier de travail dans l'ordinateur où est installé le logiciel R. On pourra par exemple créer ce dossier dans la racine « **C** » et le nommer « **outils pour la prévision des débits** » ;

b- copier dans un fichier Bloc-notes vide, le programme de prédiction « C.54 » se trouvant en Annexe du mémoire. Enregistrer ensuite ce fichier dans le dossier de travail précédemment créé en le nommant par exemple « **prog.pred** » ;

c- on suppose maintenant que les débits naturels journaliers du B.V.I. sont enregistrés dans l'ordinateur sur la forme d'un tableau de 365 lignes (ces lignes sont numérotés de 1 à 365 et représentent les différents jours d'une année comptés à partir du 1er Juillet). Ainsi, les valeurs d'une même colonne correspondent aux débits des différents jours d'une année donnée. Sélectionner sur un tel tableau de débits naturels journaliers, ceux correspondants aux années allant du 1er Juillet 1988 au 30 Juin de l'année qui précède celle que l'on désire prédire les débits. On copie cette sélection puis, on la colle dans un fichier Bloc-notes vide. Ensuite, on enregistre ce fichier dans le dossier de travail précédemment créé en le nommant par exemple « **débits journaliers du bvi** ».

3- Importation du programme de prédiction et du tableau des débits dans la console de R

a- Lancer le logiciel R en double cliquant par exemple sur son icône ;

b- changer le répertoire courant de R au dossier de travail précédemment créé en exécutant la commande suivante :

setwd("indiquer le chemin d'accès au dossier de travail créé") qui est par exemple ceci :

setwd("C:/outils pour la prévision des débits.txt") dans le cas où les noms proposés ont été utilisés ; puis, valider cette instruction en cliquant sur la touche du clavier « **Entrée** » ;

c- importer le programme de prévision dans la console de R en exécutant la commande suivante :

source("nom du fichier contenant le programme de prédiction.txt") qui est par exemple ceci : **source("prog.pred.txt")** dans le cas où le nom du fichier contenant ce programme de prédiction proposé a été utilisé ; puis, valider cette instruction en cliquant sur la touche du clavier « **Entrée** » ;

d- importer le tableau des débits dans la console de R en exécutant la commande suivante :

bvi=read.table("nom du fichier contenant les débits.txt") qui est par exemple ceci :

bvi=read.table("débits journaliers du bvi.txt") dans le cas où le nom du fichier contenant ce tableau des débits proposé a été utilisé ; puis, valider cette instruction en cliquant sur la touche du clavier « **Entrée** ».

4- Exécution de la prévision des débits naturels journaliers du B.V.I.

a- Effectuer la prévision en utilisant la commande suivante :

programme.prédiction(bvi) ; puis, valider cette instruction en cliquant sur la touche du clavier « **Entrée** ».

NB : Le résultat s'affiche sur 2 colonnes. La 1ère colonne contient les chiffres allant de 1 à 365 qui représentent les numéros des jours d'une année comptés à partir du 1er Juillet. La 2ème colonne contient les valeurs des prévisions des débits naturels associés aux différents jours de l'année dont les numéros sont indiqués par la 1ère colonne.

b- Importer les prévisions dans le logiciel Excel en exécutant les instructions suivantes :

i) sélectionner et copier les 2 colonnes obtenues comme résultat de la section 4.a ci-dessus ;

ii) ouvrir le logiciel Excel ;

iii) coller la copie dans une feuille Excel.

NB : Les 2 colonnes copiées dans R s'affichent dans Excel à l'intérieur d'une seule colonne. Afin de les faire apparaître aussi sur 2 colonnes dans Excel, il faut suivre la procédure suivante :

- cliquer sur la bulle qui s'affiche à l'extrémité inférieure droite de la colonne à l'intérieur de laquelle on a coller la copie dans Excel ;

- cocher la case : **Utiliser l'Assistance Importation de Texte…,** cette fenêtre s'ouvre ;

- cocher la case : **Délimiter**, puis cliquer sur **Suivant** ;

- cocher la case : **Espace**, puis sur **Terminer**.

NB : Les données s'affichent alors sur 2 colonnes distinctes. Les numéros des jours sont alors séparés des valeurs des débits prédits. On pourra supprimer la colonne contenant les numéros des jours d'une année et exploiter la colonne contenant les valeurs des prévisions des débits.

CHAPITRE 1

CONTEXTE ET PROBLEMATIQUE

1.1 Présentations des activités de AES-SONEL

Cette partie est inspirée de [5]. AES signifie: All energy service. AES-SONEL est une entreprise dont les principales activités sont :

- La production de l'énergie électrique ;
- Le transport et la gestion du réseau de transport de cette énergie ;
- La distribution et la vente de l'énergie électrique.

Nous présentons dans la suite les activités de production et du transport de l'énergie électrique.

1.1.1 Activités de production de l'énergie électrique

La production d'énergie électrique au Cameroun par AES-SONEL est assurée à la fois par un *système de production hydraulique* et un *système de production thermique*.

La production d'énergie électrique d'origine hydraulique est effectuée à partir des barrages hydroélectriques. Elle représente l'essentiel de la production d'électricité de AES-SONEL.

Le parc de production hydraulique comprend :

- les *centrales hydroélectriques* de Song-Loulou, Edéa et Lagdo, respectivement d'une puissance de 400, 265 et 72 MW ;
- les *barrages réservoirs* de régulation du débit du fleuve Sanaga. Il s'agit des barrages de Mbakaou, Bamendjin, Mapé et Lagdo.

Le système de production thermique fonctionne à base des combustibles tels que le fuel lourd et le fuel léger. Il est constitué :

- des centrales thermiques d'Oyomabang, Bassa, Logbaba, Bafoussam, Limbé et de Djamboutou ;
- de 31 centrales thermiques au diesel, isolées.

1.1.2 Activités de transport de l'énergie électrique

L'activité de transport est d'évacuer, vers les centres de distribution ou de consommation, l'énergie électrique produite par les centrales. Cette évacuation est assurée par un ensemble de câbles et de transformateurs connectés appelé *réseau de transport interconnecté*. Le transport de l'énergie électrique produite est assuré par deux réseaux de transport interconnectés, le réseau Sud et le réseau Nord.

- **Le Réseau interconnecté Sud (RIS)**

Le réseau de transport interconnecté Sud (RIS) permet d'évacuer vers les centres de consommation l'énergie produite par les centrales hydrauliques de Song-Loulou et d'Edéa, et les centrales thermiques d'Oyomabang, Bassa, Logbaba, Limbé et Bafoussam. Il s'étend sur 6 provinces : Centre, Sud, Littoral, Ouest, Nord-Ouest et Sud-Ouest.

- **Le Réseau Interconnecté Nord (RIN)**

Le réseau Nord est alimenté par la centrale hydroélectrique de Lagdo et la centrale thermique de Djamboutou. Il s'étend sur 3 provinces : l'Adamaoua, le Nord et l'Extrême-Nord.

On peut aussi noter que la province de l'Est, quant à elle, est alimentée par une centrale thermique isolée.

Intéressons nous, maintenant, à la production d'énergie électrique d'origine hydraulique permettant d'alimenter le réseau Sud. Pour une bonne compréhension de la production de cette énergie, il est nécessaire de connaître le bassin versant de la Sanaga ainsi que ses aménagements hydrauliques.

1.2 Présentation du bassin versant de la Sanaga

Ces informations sont issues de [9]. La Sanaga est le plus grand fleuve du Cameroun. Le bassin de la Sanaga est dans sa quasi-totalité sur le territoire Camerounais ; une infime partie ($200\,km^2$), à l'Est, se trouve en République CENTRAFRICAINE (R.C.A).

Il s'étend de la latitude 3°32' N à la latitude 7°22' N. Son point le plus Occidental se trouve sur le méridien 9°45' E ; vers l'Est, il atteint le méridien 14°57' E.

Sa superficie est d'environ $133000\,km^2$, soit plus de 25% de la superficie totale du pays. On distingue dans le cours de la Sanaga 3 grandes divisions :

- la Sanaga supérieure représentée par le Djerem ;
- la moyenne Sanaga représentée par le Mbam ;

- la Sanaga inférieure : c'est la partie du fleuve située après les chutes d'Edéa et se dirigeant vers l'océan Atlantique.

Les informations des parties 1.3, 1.4 et 1.5 sont issues du document [3].

1.3 Aménagements hydroélectriques de la Sanaga

Les aménagements hydrauliques du bassin de la Sanaga sont de 2 types :

- les ouvrages de production de l'énergie hydroélectrique ;
- les ouvrages de régularisation du débit du fleuve.

1.3.1 Les ouvrages de production d'énergie électrique

A l'heure actuelle, il y a deux sites de production situés sur le bras inférieur de la Sanaga :

- *la centrale d'Edéa* : c'est historiquement le premier site hydroélectrique, les travaux de construction ont débuté en 1933 et se sont achevés en 1958. La puissance installée a progressivement atteint la valeur de 263 MW ;
- *la centrale de Song Loulou* : cette usine, situé à 55 km en amont d'Edéa, a été équipée en deux phases (1980 et 1988) pour atteindre une puissance installée de 384 MW.

Ces deux usines fonctionnent au fil de l'eau ; ainsi toute diminution temporaire du débit du fleuve au dessous des besoins des groupes en service se traduit par une baisse immédiate de la puissance disponible.

1.3.2 Les ouvrages de régularisation du débit de la Sanaga

Le débit naturel de la Sanaga à Edéa présente de grandes variations. Il peut dépasser les 7000 m^3/s en période de crues pour descendre sous 250 m^3/s en période de basses eaux. Ce régime hydraulique irrégulier a rendu nécessaire la construction de grands barrages réservoirs d'eau en amont des usines, permettant la régularisation des débits du fleuve. En période de crue, « l'excédent » de débit permet de remplir les barrages réservoirs. Puis, en période de basses eaux, des lâchers d'eau sont effectués aux barrages réservoirs pour augmenter progressivement le débit du fleuve.

Le bassin de la Sanaga est actuellement équipé de trois barrages réservoirs situés en amont des sites de production :

- le *barrage de Mbakaou* : mis en service en 1970 sur le cours du Djerem, ce réservoir a une capacité de 2.6 milliards de mètres cubes et permet de régulariser les apports du versant sud de l'Adamaoua. La durée de propagation des éclusées entre Mbakaou et Song Loulou est de 7 jours ;

- le *barrage de Bamendjin* : mis en service en 1974 sur le cours du Noun, ce réservoir a une capacité de 1.8 milliards de mètres cubes. La durée de propagation des éclusées entre Bamendjin et Song Loulou est de 5 jours ;
- le *barrage de Mapé* : mis en service en 1987 sur le cours de la Mapé, ce réservoir a une capacité de 3,2 milliards de mètres cubes. La durée de propagation des éclusées entre Mapé et Song Loulou est de 5 jours.

1.3.3 Les stations hydrologiques du B.V.I. de la Sanaga

Environ 80% de l'énergie électrique camerounaise est produite à partir des usines hydroélectriques. La gestion de la production d'énergie électrique d'origine hydraulique nécessite une bonne connaissance des débits des cours d'eau qui alimentent ces usines. Dans le but de suivre l'évolution effective des lâchers d'eau effectués au niveau des barrages réservoirs, un réseau de stations de jaugeages a été construit en aval de ces barrages réservoirs pour le contrôle des débits.

La figure 1.2 de la page 17 présente les stations hydrologiques du B.V.I. de la Sanaga. Notons que la station hydrologique de Songmbengué est le dernier point de contrôle des débits. Elle est située à 4 heures de la centrale hydroélectrique de Song-Loulou. Ainsi le débit mesuré à cette station est exactement celui de l'eau qui alimentera cette centrale.

Nous présentons sur la figure suivante le système de production hydroélectrique du bassin de la Sanaga.

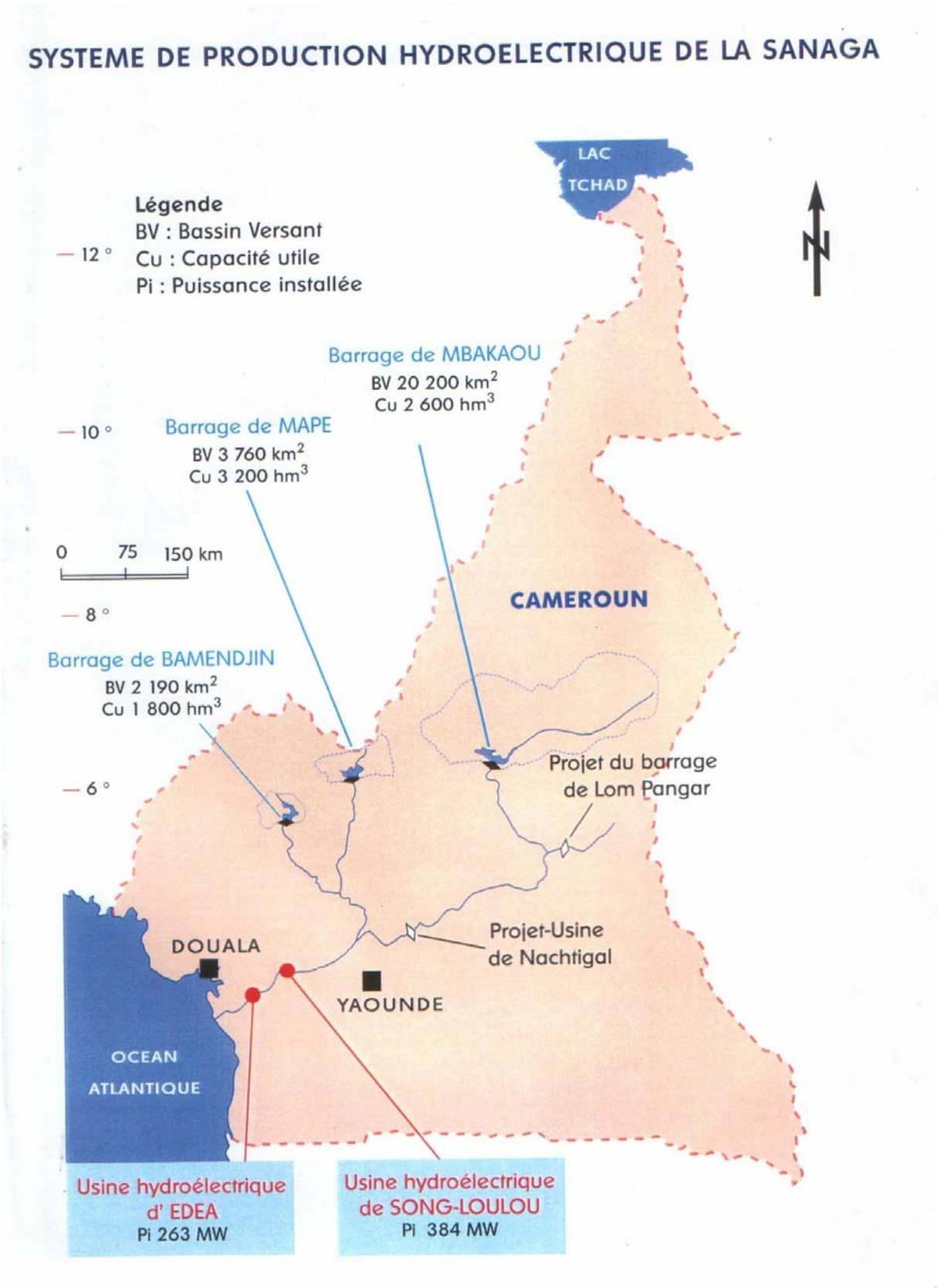


FIG. 1.1 : Système de production hydroélectrique de la Sanag

1.4 Régularisation de la Sanaga

L'ensemble des trois barrages réservoirs présentés dans la section 1.4.2 permet de réguler le débit de la Sanaga. Le débit naturel de la rivière est donc remplacé par un débit artificiel répondant à une règle de gestion bien précise définie par le gestionnaire, en l'occurrence AES-SONEL.

Ainsi, suivant le débit disponible à l'entrée des ouvrages de production, le niveau de la demande en énergie électrique et l'état des stocks d'eau dans l'ensemble de ces trois barrages réservoirs, des lâchers d'eau sont effectués pour augmenter le débit au niveau des usines de production en vu de répondre à la demande en énergie.

Cependant, les stocks d'eau de ces trois barrages réservoirs ne sont pas gérés selon les mêmes consignes. Le barrage de Mbakaou a une capacité très petite devant les apports moyens de la rivière en amont du site, c'est-à-dire que le volume total d'eau récolté par son bassin versant au cours d'une année est très supérieur à la quantité d'eau maximale que le barrage peut stocker. Ainsi, la mise en réserve au moment de la crue annuelle va de pair avec l'évacuation d'une grande partie de celle-ci et le rôle régulateur est somme toute assez modeste sur le bassin aval. Comme il est assuré de se remplir chaque année au cours de la saison humide, l'intégralité de son stock peut-être utilisée en cours de la saison sèche de la même année sans prendre le risque de compromettre le remplissage du réservoir au cours de la prochaine saison.

Par contre, les barrages réservoirs de la Mapé et de Bamendjin ont des capacités de stockage à peu près égales aux apports moyens de leurs bassins versants respectifs. La conséquence est qu'ils ne doivent pas être vidés complètement en une seule saison d'étiage, car leur remplissage pendant la saison humide de l'année suivante n'est pas acquis. En effet, pour les remplir à nouveau, il faudrait que l'année suivante soit une année « moyenne » au sens hydrologique. Ils assurent donc une régularisation inter annuelle, c'est-à-dire que les apports d'une année peuvent être restitués les années suivantes.

Une gestion courante de ce type de parc d'aménagement consisterait donc, pendant la période sèche, à utiliser prioritairement le barrage réservoir de Mbakaou jusqu'à vidange complète de la retenue, puis à utiliser conjointement ceux de Mapé et de Bamendjin, tout en y conservant un stock minimal suffisant, garantissant leur remplissage l'année suivante.

Dans le cas de la Sanaga, cette gestion est très délicate. En effet, les centrales hydroélectriques sont très éloignées des barrages réservoirs de régularisation. Le temps mis par un volume d'eau lâché à Mbakaou pour parvenir aux sites de production est d'environ une

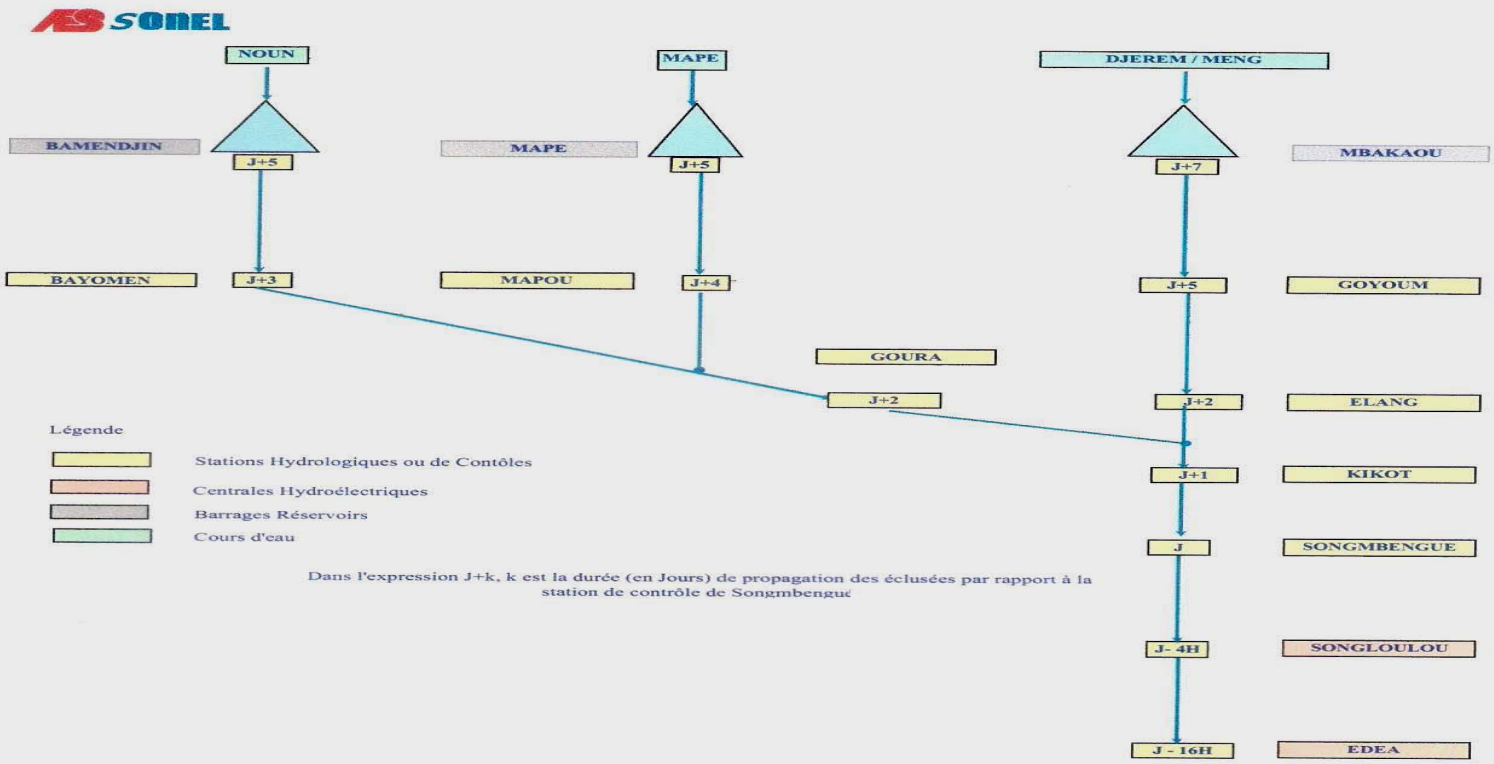
semaine. De plus, le « train d'eau » libéré par le réservoir a une tendance à s'aplatir et s'amortir le long du lit naturel de la rivière.

Enfin, les précipitations perturbent et compliquent cette gestion. En effet, de fortes précipitations, survenant après un important lâcher, auraient pour conséquence première de provoquer un apport supplémentaire d'eau sur le bassin versant intermédiaire de la Sanaga. Cet apport venant s'ajouter au lâcher aurait pour conséquence de saturer les capacités de production aux usines. Une partie des apports serait donc déversée sans être turbinée. Ainsi, de l'eau stockée aurait été lâchée inutilement et, à terme, les barrages réservoirs pourraient se vider prématurément avant la fin de la saison des basses eaux. A l'inverse, des lâchers trop faibles sans pluies importantes provoqueraient une baisse significative de la puissance disponible.

Pour assurer une bonne exploitation de l'eau des retenues, on doit prévoir à l'avance le débit naturel de la Sanaga aux durées de transit des éclusées pour lâcher des quantités d'eau juste nécessaires pour maintenir le débit aux usines de production à la valeur requise pour assurer la puissance électrique appelée aux bornes des alternateurs.

La régularisation de la Sanaga par des barrages réservoirs très éloignés des ouvrages de production est à peu près unique en son genre dans le monde pour les aménagements d'une telle ampleur.

FIG 1.2 : Stations hydrologiques du B.V.I.



Une fois le contexte présenté, exposons maintenant l'origine de l'objet de notre étude.

1.5 Problématique Générale

Les usines hydroélectriques d'Edéa et de Song-Loulou fonctionnent au fil de l'eau, ainsi toute diminution temporaire de débit au dessous des besoins des groupes en service se traduit par une baisse immédiate de la puissance électrique disponible. Le régime hydrologique de la Sanaga, très varié, entraînait une production d'énergie en dents de scie complètement liée à la merci de la nature.

Afin d'assurer la production d'énergie électrique juste nécessaire pour satisfaire la demande tout en réduisant au maximum les pertes en eaux, il est nécessaire de connaître cinq jours à l'avance le débit naturel de la Sanaga à la dernière station de contrôle de Songmbengué. En effet, cela permettra de planifier de façon optimale les quantités d'eau à lâcher des barrages réservoirs de Mapé et de Bamendjin qui permettront de compenser ce débit naturel. Ceci, si l'on suppose que le débit de l'eau lâchée du barrage de Mbakaou est constant.

Depuis quelques années, la production d'énergie électrique est en croissance. Le poste de la dépense en volume de combustibles est passé au second rang des dépenses globales de l'entreprise. Ce qui implique une planification à moyen terme (12 mois) de cette dépense. Afin de mieux cadrer cette source de dépense en volume de combustibles, une prévision des débits naturels journaliers à moyen terme est indispensable.

CHAPITRE 2

IDENTIFICATION ET DESCRIPTION DES DONNEES

2.1 Identification et obtention des données

Nous identifierons les apports naturels aux débits naturels**.** L'apport naturel n'est pas mesuré directement et il est obtenu à partir de l'équation du bilan hydrique qui fait intervenir plusieurs composantes (données mesurées ou déduites d'équations approximatives) reposant sur différentes hypothèses concernant les comportements hydrauliques. Ces valeurs peuvent être entachées d'erreurs importantes pouvant se traduire par des variations excessives, irréalistes et des biais systématiques dans les séries d'apports.

Vu la complexité des méthodes d'obtention des débits naturels reconstitués, nous analyserons ces derniers sans tenir compte de la correction des observations qui peuvent être considérées comme aberrantes.

Comme la prédiction des débits d'un cours d'eau ne demande pas la connaissance exacte de tous les processus de génération des écoulements sur le bassin versant, les données que nous utiliserons pour réaliser cette étude sont les débits naturels journaliers du bassin versant intermédiaire de la Sanaga à la station de Songmbengué, calculés sur la période allant du 1er Juillet 1988 au 30 Juin 2005. Le choix de cette station est motivé par le fait qu'il s'agit du dernier point de contrôle du débit du fleuve provenant du bassin versant de la Sanaga et se dirigeant vers les ouvrages de production d'énergie électrique. Le choix de 1988 comme date de début de la chronique est guidé par le fait qu'il s'agit de l'année à partir de laquelle le calcul de ces débits naturels à la dite station tient compte de l'influence simultanée des trois barrages réservoirs situés en amont.

Ces débits naturels proviennent des apports naturels du bassin versant de la Sanaga, et sont en réalité les débits naturels reconstitués. C'est-à-dire les débits qu'aurait parcouru le fleuve si celui-ci n'avait pas été équipé des barrages réservoirs.

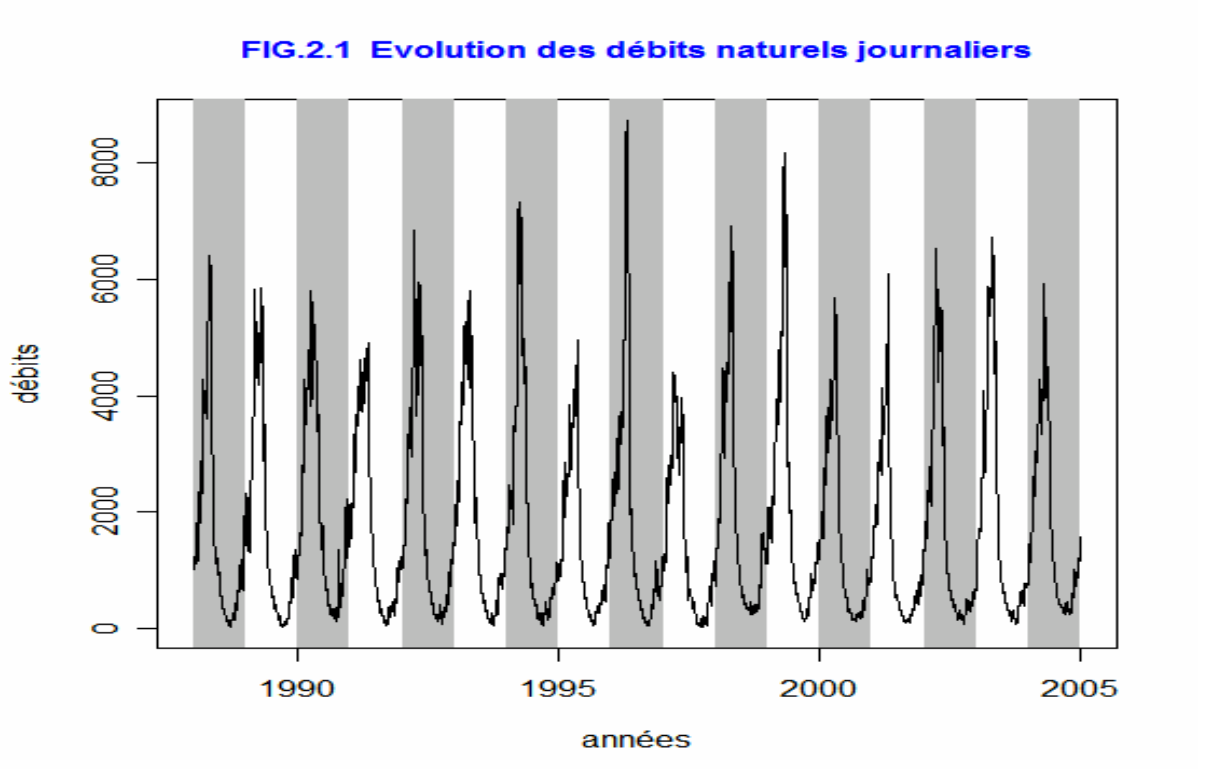
Ces données sont stockées dans un tableau au format Excel de 365 lignes qui représentent les jours d'une année et de 17 colonnes qui représentent les années d'observations ou de calculs. La chronique s'étend du 1er Juillet 1988 au 30 Juin 2005.

2.2 Description des données

Tout au long de ce travail, nous utiliserons le logiciel R (version non commerciale de S-plus), téléchargeable sur le site internet http://lib.stat.cmu.edu/R/CRAN/ aussi bien pour faire les calculs numériques, les graphiques que pour faire les prévisions et les simulations. Tous les commandes R utilisées sont stockées en annexe. Notons aussi que la première valeur des hydrogrammes de chapitre correspond à la valeur du débit du 1er Juillet.

La figure 2.1 présente l'évolution des débits naturels journaliers reconstitués du B.V.I. de la Sanaga à la station hydrologique de Songmbengué sur toute la période d'étude. C'est-à-dire du 1er Juillet 1988 au 30 Juin 2005. (Voir commandes C.1 pour son obtention).

FIG.2.1 Evolution des débits naturels journaliers



On remarque que la tendance de cette chronique est presque inexistante ou du moins, varie très peu. Donc, cette chronique paraît essentiellement composée des variations saisonnières.

Les trois figures suivantes présentent les évolutions journalières des débits naturels de la Sanaga à la station de contrôle de Songmbengué par an. Notons que les débits d'une année vont du 1er Juillet au 30 Juin de l'année suivante. (Voir commandes C.2, C.3, C.4 pour leur obtention).

FIG.2.2 Evolution journalière des débits naturels des années 1988 à 1994

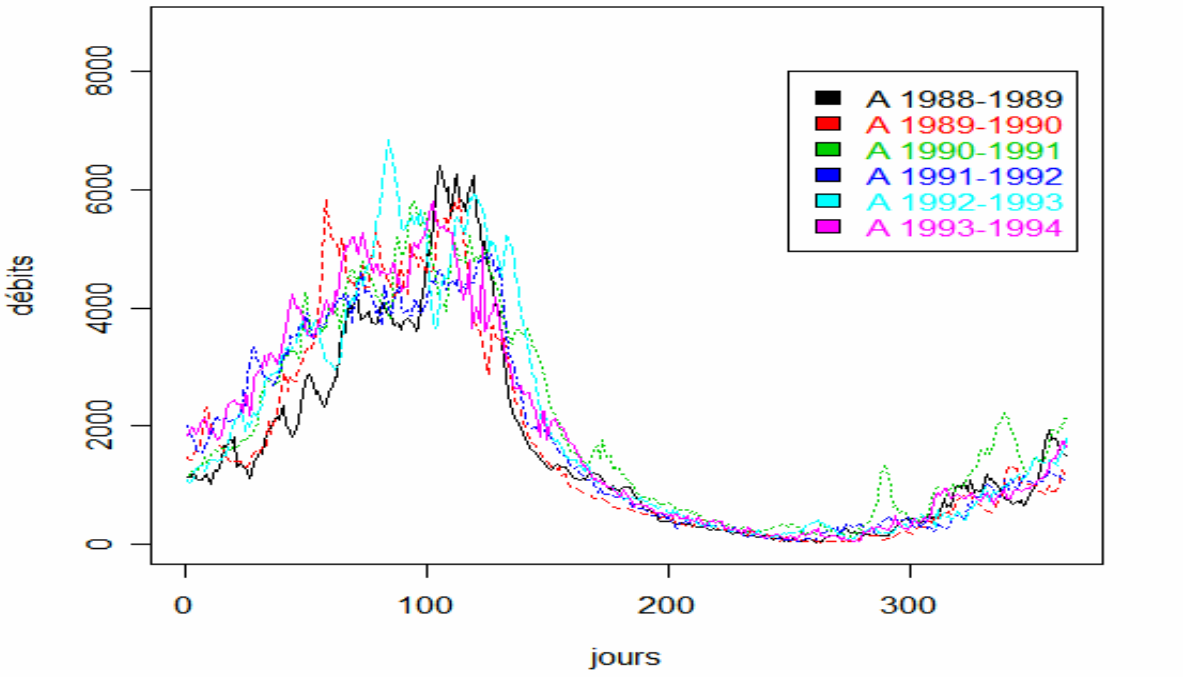
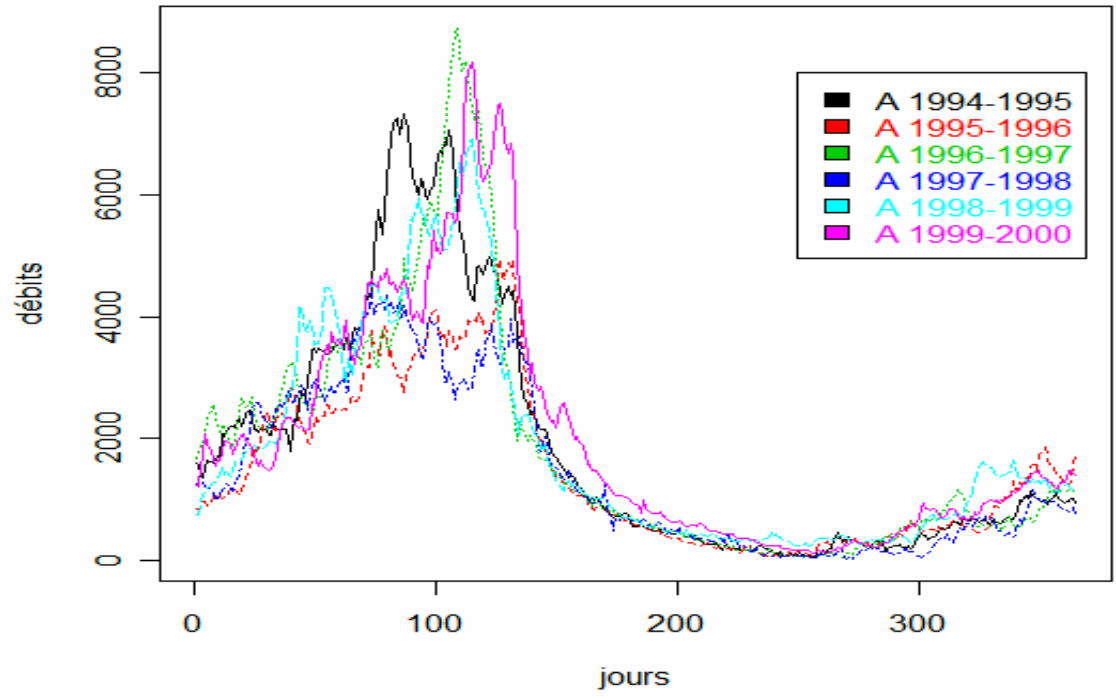
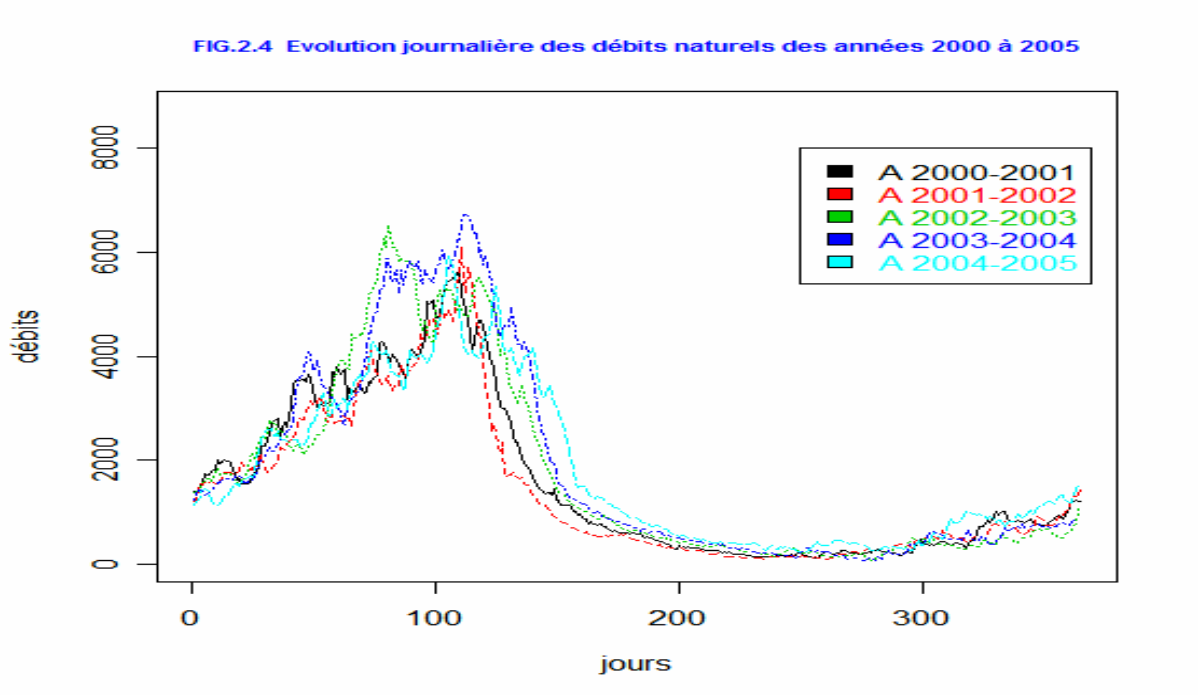


FIG.2.3 Evolution journalière des débits naturels des années 1994 à 2000





Nous remarquons, des figures ci-dessus, que le débit naturel de la Sanaga, à Songmbengué, a une évolution cyclique dont la période est de 12 mois ; chaque cycle est formé d'une succession de 2 phases :

- une phase de crue ou croissance des débits qui s'étend du mois d'Avril au mois d'Octobre où le débit maximal est atteint. Cette valeur maximale du débit est généralement supérieure à 6000 m^3 / s .
- une phase d'étiage ou de décroissance des débits qui s'étend du mois de Novembre au mois de Mars de l'année suivante où le débit minimal est atteint. Cette valeur est généralement inférieure à 100 m^3 / s .

Ainsi, l'année hydrologique ici est la période allant du mois d'Avril au moi de Mars de l'année suivante.

Les courbes présentant les évolutions journalières des débits naturels du B.V.I. de la Sanaga à Songmbengué suivant un découpage mensuel sont présentées sur les figures 2.5 à 2.40 dans l'annexe. Le tableau contenant les débits naturels journaliers du B.V.I. de la Sanaga à Songmbengué ainsi que les valeurs des paramètres de dispersion et de tendance centrale correspondants à ces débits, est présenté dans l'annexe (tableau 2.1)**.** Le tableau contenant les débits moyens mensuels correspondants à ces débits naturels journaliers, est présenté dans l'annexe (tableau 2.2)**.**

CHAPITRE 3

OUTILS STATISTIQUES

3.1 Modèles déterministes

Définition 3.1 Une série chronologique $\{Y_t, t \in T\}$ est une suite d'observations indexées par un ensemble ordonné $T = \{t_1, t_2, ..., t_n\}$.

On supposera, dans toute la suite, que les dates sont équidistantes et donc, nous adopterons la notation simplifiée pour l'ensemble d'indices $T = \{1, 2, ..., n\}$. Ainsi, la série s'écrira $\{Y_t, t = 1, 2, ..., n\}$.

On considère qu'une série $\{Y_t\}$ est la résultante de différentes composantes fondamentales :

- **La tendance** (ou **trend**) $\{C_t\}$ représente l'évolution à long terme de la série ;
- **La composante saisonnière** ou **saisonnalité** $\{S_t\}$ correspond à un phénomène qui se répète à intervalles de temps réguliers (phénomène périodique). En général, c'est un phénomène saisonnier, d'où le terme de variations saisonnières. La composante saisonnière est donc périodique de période p et il suffit de connaître ses p premières valeurs $S_1, S_2, ..., S_p$ (par périodicité, on a $S_t = S_{t+p}$ pour tout t).
- **La composante résiduelle** ou **bruit** ou **résidu** $\{\varepsilon_t\}$: représente les fluctuations irrégulières, en général de faible intensité, mais de nature aléatoire.

Les composantes ci-dessus peuvent se combiner selon différents modèles.

a- Modèle additif

$$Y_t = C_t + S_t + \varepsilon_t, \text{ avec } t = 1, 2, ..., n,$$

$$\text{Hypothèses : } \sum_{j=1}^{p} S_j = 0 \text{ et } \sum_{t=1}^{n} \varepsilon_t = 0.$$

b- Modèle multiplicatif

$$Y_t = C_t \times S_t \times \varepsilon_t \quad \text{avec} \quad t = 1,2,...,n,$$

$$\text{Hypothèses : } \sum_{j=1}^{p} S_j = p \quad \text{et} \quad \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t = 1.$$

Pour effectuer l'analyse d'une série chronologique, on essaie d'abord de déterminer si les composantes de cette série peuvent être combinées selon un modèle additif ou un modèle multiplicatif. Afin de faire cette distinction, nous présentons une méthode d'identification décrite dans [7], en supposant que la série a une périodicité annuelle.

Méthode d'identification du type de schéma

1- On calcule d'abord les moyennes et les écarts-types pour chacune des périodes considérées. C'est-à-dire, par exemple pour une chronique ayant une périodicité annuelle, on calcule les moyennes et les écarts-types des observations pour chacune des années de cette chronique.

2- On calcule ensuite la droite des moindres carrés $\sigma = a\bar{x} + b$. C'est-à-dire, on calcule la droite de régression de ces écarts-types en fonction de ces moyennes.

3- Enfin, si a est significativement non différent de 0 à un seuil que l'on s'est fixé (par exemple 5%), nous pouvons conclure que le modèle de composition est additif, sinon, le modèle de composition est multiplicatif.

Définition 3.2 Une série des **moyennes mobiles d'ordre k,** notée $MM(k)_t$, est la série des moyennes de k observations consécutives et elle prend ses valeurs aux dates moyennes correspondantes. Plus précisément, on calcule les moyennes de k termes consécutifs :

1. pour les dates : $\frac{t_1 + t_2 + ... + t_k}{k}$, puis $\frac{t_2 + t_3 + ... + t_{k+1}}{k}$, …, jusqu'à
   $\frac{t_{n-k} + t_{n-k+1} + ... + t_n}{k}$;
2. et pour la variable d'intérêt :
   $\frac{Y_1 + Y_2 + ... + Y_k}{k}$, puis $\frac{Y_2 + Y_3 + ... + Y_{k+1}}{k}$, …, jusqu'à
   $\frac{Y_{n-k} + Y_{n-k+1} + ... + Y_n}{k}$.

Remarque 3.1 Si k est impair : $k = 2m + 1$, la série moyenne mobile est calculée aux mêmes instants que les observations initiales. Ainsi, les dates de la série des moyennes

mobiles sont des entiers naturels. En revanche, lorsque k est pair : $k = 2m$, la série moyenne mobile est calculée entre les dates des observations de la série initiale. Ainsi, les dates de la série des moyennes mobiles ne sont pas des entiers naturels.

Ainsi, une moyenne mobile d'ordre pair se calcule à des dates qui ne coïncident pas avec les dates des observations. Si l'on veut comparer la série des moyennes mobiles avec la série initiale, on a besoin d'avoir des valeurs pour les mêmes dates d'observation. On définit les moyennes mobiles centrées pour pallier cet inconvénient des moyennes mobiles d'ordre pair.

Définition 3.3 En gardant les notations ci-dessus, on définit la série des **moyennes mobiles centrées** d'ordre $k = 2m$, notée $MMC(k)_t$, par :

$$MMC(k)_t = \frac{0.5 \times Y_{t-m} + ... + Y_t + ... + 0.5 \times Y_{t+m}}{2m},\ t = m+1,...,n-m;$$

et d'ordre $k = 2m+1$, par :

$$MMC(k)_t = \frac{Y_{t-m} + ... + Y_t + ... + Y_{t+m}}{2m+1},\ t = m+1,...,n-m\ ;$$

n est le nombre total des observations de la série initiale.

Remarque 3.2 Si $k = 2m$ ou $k = 2m+1$, m observations sont perdues à chaque extrémité de la série des moyennes mobiles centrées d'ordre k. Ainsi, la série initiale et la série des moyennes mobiles centrées d'ordre k n'ont pas la même longueur pour $k > 1$.

Proposition 3.1 Si une série chronologique à une composante saisonnière de période p, alors une moyenne mobile d'ordre p absorbe cette composante saisonnière, c'est-à-dire l'annule. (Cf [1]).

3.2 Un algorithme de modélisation d'une chronique

Nous présentons dans cette partie un algorithme permettant de caractériser une chronique par un modèle déterministe. Pour simplifier l'exposé, nous présentons cette démarche en l'appliquant à une chronique dont la fréquence des observations est mensuelle. Cette méthodologie sera appliquée au chapitre 4 pour modéliser une chronique dont la fréquence des observations est journalière.

Soit une série chronologique $(Y_t)_{t=1,...,np} = (Y_{i,j})_{\substack{i=1,...,n \\ j=1,...,p}}$, t est le nombre de mois à partir de la date 0 ; i est le numéro de l'année ; j est le numéro du mois dans l'année i.

1. On estime la tendance en éliminant la composante saisonnière à l'aide des moyennes mobiles centrées dont l'ordre est la période de la saisonnalité.

$$C_t = MMC(k)_t,\ t = p(i-1)+j,\ i = 1,2,...,n,\ j = 1,2,...,p.$$

2. On identifie le modèle de composition : schéma additif ou multiplicatif.

a- Cas d'un modèle additif :

On exclut les valeurs extrêmes perdues lors du calcul de la série des moyennes mobiles centrées. Puis,

- on calcule les données sans tendance (ou différences saisonnières) $Y_t - C_t$;
- on calcule la moyenne des données sans tendance du mois j sur les n années, ceci pour chacun des mois. Ce sont les coefficients saisonniers.

$$\text{D'où } S_j = \frac{1}{n}\sum_{i=1}^{n}\left(Y_{i,j} - C_{i,j}\right);$$

- on calcule la moyenne des S_j : $\overline{S} = \frac{1}{p}\sum_{j=1}^{p} S_j$.

Si $\overline{S} \neq 0$, on corrige les S_j : $S'_j = S_j - \overline{S}$.

b- Cas d'un modèle mulptiplicatif :

- On calcule les données sans tendance (ou rapports saisonniers) : $\frac{Y_t}{C_t}$;
- on calcule la moyenne des données sans tendance du mois j sur les n années, ceci pour chacun des mois,

$$S_j = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_{i,j}}{C_{i,j}};$$

- on calcule la moyenne des S_j : $\overline{S} = \frac{1}{p}\sum_{j=1}^{p} S_j$.

Si $\overline{S} \neq 1$, on corrige les S_j : $S'_j = \frac{S_j}{\overline{S}}$.

On obtient à la fin de cette 2ième étape, la série des variations saisonnières :

Pour tout i, $S_{i,j} = S'_j$ ceci pour tous les mois j.

3. On calcule la série corrigée des variations saisonnières (CVS)

a- Cas d'un modèle additif :

$$D_{i,j} = Y_{i,j} - S_{i,j} = Y_{i,j} - S'_j .$$

b- Cas d'un modèle multiplicatif :

$$D_{i,j} = \frac{Y_{i,j}}{S_{i,j}} = \frac{Y_{i,j}}{S'_j} .$$

On peut réévaluer la tendance à l'issue de cette 3ième étape par ajustement global de la série corrigée des variations saisonnières (CVS).

4. On calcule la série ajustée

a- Cas d'un modèle additif :

$$\hat{Y}_t = C_t + S_t \qquad \text{ou} \quad \hat{Y}_{i,j} = C_{i,j} + S'_j .$$

b- Cas d'un modèle multiplicatif :

$$\hat{Y}_t = C_t \times S_t \quad \text{ou} \quad \hat{Y}_{i,j} = C_{i,j} \times S'_j .$$

5. Calcul des variations accidentelles ou résiduelles

a- Cas d'un modèle additif :

$$\varepsilon_t = Y_t - \hat{Y}_t .$$

b- Cas d'un modèle multiplicatif :

$$\varepsilon_t = \frac{Y_t}{\hat{Y}_t} .$$

On a ainsi décomposé la série chronologique (Y_t) en 3 composantes : sa tendance C_t, ses variations saisonnières S_t et ses variations accidentelles ε_t qui se combinent de la manière suivante :

a. pour le schéma additif : $Y_t = C_t + S_t + \varepsilon_t$;

b. pour le schéma multiplicatif : $Y_t = C_t \times S_t \times \varepsilon_t$.

Le modèle étant retenu, on peut faire des prévisions très facilement. On prévoit la tendance en calculant $C_{np+1}, C_{np+2}, \ldots$, puis, selon le modèle de composition, on ajoute ou on multiplie par le coefficient saisonnier du mois correspondant.

3.3 Modèles stochastiques

Soit $(\Omega, \mathrm{A}, \mathrm{P})$ un espace probabilisé, (T, Γ) et (Ω', A') deux espaces mesurables.

Définition 3.4 Un **processus stochastique** est une application X définie sur $\Omega \times \mathrm{T}$, à valeurs dans Ω', associant au couple (ω, t) la réalisation $X(\omega, t)$, encore notée $X_t(\omega)$, et tel que pour t fixé appartenant à T, X est une variable aléatoire (v.a.) sur (Ω, A).

Par extension, on écrira un processus sous la forme d'une famille de v.a. indicées par t, notée $(X_t, t \in \mathrm{T})$ ou, plus simplement (X_t). La loi du processus est l'image P^X de P par X.

Lorsque $\Omega' = \mathbb{R}$, le processus est dit unidimentionnel ou univarié. Lorsque $\mathrm{T} = \mathbb{Z}$, le processus est dit en temps discret.

Nous considérerons dans la suite les processus stochastiques univariés et à temps discret.

Définitions 3.5 Les **modèles de simulation** sont des modèles qui ne cherchent pas à reproduire une partie de la réalisation d'un processus, mais à générer des scénarios « possibles » d'un processus dont on connait une réalisation. Leur principe est d'utiliser les générateurs de variables aléatoires qui devront respecter la structure statistique des processus à reproduire.

Définition 3.6 Les **modèles de prévision** sont les modèles qui cherchent à un instant donné t, à « prévoir » pour les instants $t + \Delta t$ le devenir d'une réalisation connue jusqu'à t.

Définition 3.7 L'**opérateur retard** B est un opérateur qui, à un processus X_t, associe le processus Y_t tel que $Y_t = BX_t = X_{t-1}$.

Définition 3.8 On dit que X_t est **strictement (ou fortement) stationnaire** si pour toute suite finie d'instants $t_1, t_2, ..., t_k$ éléments de $\mathbb{Z}$ et tout entier $r \in \mathbb{Z}$, les lois jointes de $\left(X_{r+t_1}, X_{r+t_2}, ..., X_{r+t_k}\right)$ et de $\left(X_{t_1}, X_{t_2}, ..., X_{t_k}\right)$ sont les mêmes (lois jointes invariantes par translation dans le temps).

Définition 3.9 Un processus X_t est **stationnaire au second ordre (ou faiblement stationnaire)** si ses moyennes et ses covariances sont invariantes par translation. C'est-à-dire :

$$\begin{cases} \mu_t = E(X_t) = \mu, \\ \gamma(t, j) = E\left(\left(X_{t+j} - \mu\right)\left(X_t - \mu\right)\right) := \gamma(j) \end{cases}, \text{ pour tout } t, j \in \mathbb{Z}.$$

Dans la suite, stationnaire signifiera stationnaire au second ordre. Supposons que X_t est un processus stationnaire au second ordre.

Définition 3.10 γ est appelé **fonction d'autocovariance.**

Remarque 3.3 $\text{var}(X_t) = \gamma(0)$, pour tout $t \in \mathbb{Z}$.

Théorème 3.1 ([1])

(i) $\gamma(j) = \gamma(-j)$, pour tout $j \in \mathbb{Z}$;

(ii) $\gamma(0) \geq 0$;

(iii) γ est une fonction définie positive, c'est-à-dire :

$$\sum_{j=1}^{n}\sum_{k=1}^{n}\gamma\left(t_k - t_j\right)z_j z_k \geq 0,\ \forall \mathrm{n} > 0,\ \forall\left(\mathrm{t_i}\right)_{i=1}^{n},\left(z_i\right)_{i=1}^{n} \in \mathbb{R}^n.$$

On pose pour tout $j \in \mathbb{Z}$, $\rho(j) = \dfrac{\gamma(j)}{\gamma(0)}$.

Définition 3.11 ρ est appelé **fonction d'autocorrélation.**

Corollaire 3.1 La fonction d'autocorrélation ρ a toutes les propriétés de la fonction d'autocovariance et satisfait la condition supplémentaire $\rho\left(0\right) = 1$.

On pose :

$$r(k) = \frac{\text{cov}(X_1 - P_{[X_2,X_3,\ldots,X_k]}(X_1)\ ,\ \mathrm{X_{k+1}} - P_{[X_2,X_3,\ldots,X_k]}(X_{k+1}))}{\text{var}(X_1 - P_{[X_2,X_3,\ldots,X_k]}(X_1))},\ \text{pour}\ k \in \mathbb{Z},$$

où $P_{[X_2,X_3,\ldots,X_k]}(X_1)$ et $P_{[X_2,X_3,\ldots,X_k]}(X_{k+1})$, sont respectivement la régression ou la projection de X_1 (respectivement de X_{k+1}) sur le sous-espace engendré par les variables aléatoires $X_2, X_3, \ldots, X_k$.

Définition 3.12 r est appelé **fonction d'autocorrélation partielle**.

Définition 3.13 Un **bruit blanc** est un processus ε_t, $t \in \mathbb{Z}$, centré tel que :

$$E\left(\varepsilon_t \varepsilon_s\right) = \sigma^2 \delta_s^t,\ \text{avec}\ \delta_s^t = \begin{cases} 1,\ \text{si}\ t = s \\ 0,\ \text{sinon} \end{cases},\ \sigma > 0.$$

Proposition 3.2 Un estimateur empirique de la fonction d'autocorrélation (**ACF**) ρ est définie par : $\hat{\rho}_k = \dfrac{\hat{\gamma}(k)}{\hat{\gamma}(0)}$

où $\hat{\gamma}_k = \begin{cases} \frac{1}{n-k}\sum_{t=1}^{n-k}\left(X_{t+k}-\overline{X}\right)\left(X_t-\overline{X}\right), & \text{si } 0 \le k \le n-1 \\ 0, & \text{si } k \ge n. \end{cases}$ et $\overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$.

Proposition 3.3 Un estimateur empirique de la fonction d'autocorrélation partielle (**PACF**) r s'obtient en résolvant le système de Yule-Walker suivant, en $(\alpha_1,...,\alpha_k)$,

$$\hat{\rho}(j) = \alpha_1\hat{\rho}(j-1) + \alpha_2\hat{\rho}(j-2) + ... + \alpha_k\hat{\rho}(j-k),\ j = 1,2,...,k\ ,$$

et en prenant $\hat{r}(k) = \alpha_k$, $k \in \mathbb{Z}$. (Cf [1]).

Définition 3.14 X_t est un processus $ARMA(p,q)$ (**autorégressif moyenne mobile)** s'il est stationnaire et vérifie : $\phi(B)X_t = \theta(B)\varepsilon_t$, $t \in \mathbb{Z}$ où ε_t est un bruit blanc,

$$\phi(B) = II - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p\ ,\quad \phi_p \neq 0 \text{ et}$$

$$\theta(B) = II + \theta_1 B + \theta_2 B^2 + ... + \theta_q B^q\ ,\quad \theta_q \neq 0\ ,$$

θ et ϕ sont des polynômes de degré q et p respectivement et donc les racines sont de module supérieur à 1 et ne sont pas communes.

Définition 3.15 Un processus **auto régressif** d'ordre p, noté $AR(p)$, est un processus $ARMA(p,q)$ avec $q = 0$.

Définition 3.16 Un processus **moyenne mobile** d'ordre q, $MA(q)$, est un processus $ARMA(p,q)$ avec $p = 0$.

Définition 3.17 Un processus X_t est **intégrée** d'ordre d si les processus $(1-B)^n X_t, n = 1,2,...,d-1$, ne sont pas asymptotiquement équivalents à un processus stationnaire, mais la série $Y_t = (1-B)^d X_t$ l'est.

Définition 3.18 X_t est un processus $ARIMA(p,d,q)$ (**autorégressif moyenne mobile intégré)** s'il vérifie une équation du type :

$\phi(B)\Delta^d X_t = \theta(B)\varepsilon_t$, $t \in \mathbb{N}$, où ε_t est un bruit blanc, $\Delta^d X_t = (1-B)^d X_t$,

$$\phi(B) = II - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p\ ,\ \phi_p \neq 0\ ,$$

$$\theta(B) = II + \theta_1 B + \theta_2 B^2 + ... + \theta_q B^q\ ,\ \theta_q \neq 0\ ,$$

θ et ϕ sont des polynômes de degré q et p respectivement et donc les racines sont de module supérieur à 1 et où les conditions initiales

$Z_{-1} = \left\{X_{-1},...,X_{-p},\varepsilon_{-1},...,\varepsilon_{-q}\right\}$, sont non corrélées avec $\varepsilon_0,\varepsilon_1,...,\varepsilon_t,...$.

Proposition 3.4 Si X_t est un processus $ARIMA(p,d,q)$, alors le processus $Y_t = (1-B)^d X_t$ est asymptotiquement un $ARMA(p,q)$. (Cf [1]).

Définition 3.19 X_t est un processus $SARIMA_S\left[(p,d,q),(P,D,Q)\right]$ c'est-à-dire un processus $ARIMA$ **saisonnier,** s'il vérifie une équation de type :

$\Delta^d \phi_p(B)\Delta_S^D \phi_P(B^S)X_t = \theta_q(B)\theta_Q(B^S)\varepsilon_t$, où ε_t est un bruit blanc ;

S est la période de la saisonnalité ; $\Delta_S = 1 - B^S$; $\phi_p, \phi_P, \theta_q, \theta_Q$ sont des polynômes de degré respectif p, P, q, Q dont les racines sont de modules supérieurs à 1.

Remarque 3.4 Les processus définis ci-dessus ont la forme générale suivante :

$\Phi^*(B)X_t = \Theta^*(B)\varepsilon_t$, où Φ^*, Θ^* sont des polynômes et ε_t est un bruit blanc.

Définition 3.20 Φ^* est appelé **polynôme autorégressif** du processus.

Proposition 3.5 Si X_t est un processus $AR(p)$, alors $r(k) = 0$ si $k > p$ et $r(p) \neq 0$.

Proposition 3.6 Si X_t est un processus $MA(q)$, alors $\rho(h) = 0$ si $|h| > q$ et $\rho(q) \neq 0$.

3.4 Méthodologie : modèle $ARIMA(p,d,q)$

On dispose des observations $x_1,...,x_T$ de $X_1,...,X_T$. Comment modéliser par un modèle $ARIMA(p,d,q)$ **?**

3.4.1 Identification à priori et estimation

<u>Première phase de l'identification</u> : choix de d

Approche empirique : l'autocorrélogramme

Si les $\hat{\rho}_T(h)$ restent proches de 1 ou décroissent lentement avec h, alors le processus est sans doute non stationnaire.

Remarque 3.5 Si l'autocorrélogramme fait penser que X_t est non stationnaire, alors on étudie l'autocorrélogramme du processus $Y_t = (1-B)X_t$, etc.…

Approche par test de racine unité : test de Phillips-Perron

L'hypothèse nulle est l'hypothèse de non stationnarité dans la série étudiée. La non stationnarité ici est caractérisée par la présence d'une racine unité dans le polynôme autorégressif du modèle. Pour plus d'informations sur ce test (cf [1] et [4]).

Remarque 3.6 Si ce test permet de ne pas rejeter l'hypothèse nulle, alors X_t est intégré d'ordre au moins 1. On effectue à nouveau ce test avec le processus $Y_t = (1-B)X_t$, etc.…

La valeur de d est celle pour laquelle le processus $(1-B)^d X_t$ est stationnaire.

Deuxième phase de l'identification : **choix de p et q**

On suppose que l'on a déjà d et on travaille éventuellement sur $Y_t = (1-B)^d X_t$. On assimile Y_t a un processus $ARMA(p,q)$. On se propose donc de déterminer p et q.

Approche empirique : on cherche ici à déterminer la valeur de P et Q telle que Y_t est un $ARMA(P,0)$ et un $ARMA(0,Q)$, ce qui est à peu près équivalent pour K grand à $r(k)=0$ et $\rho(h)=0$ pour $P+1 \le k \le K$ et $Q+1 \le h \le K$. (K est le nombre maximal des autocorrélations que l'on désire calculer).

Une fois P et Q déterminés, Box et Jenkins proposent en général de traiter séparément toutes les possibilités des couples (p,q) majorés par (P,Q).

Estimation

A l'issue des phases précédentes, on a choisi d et divers couples (p,q) compatibles avec les données. Le modèle s'écrit :

$$\phi(B)\Delta^d X_t = \theta(B)\varepsilon_t,$$

où ε_t est un bruit blanc de variance σ^2 et $\phi(B) = II - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$, $\theta(B) = II + \theta_1 B + \theta_2 B^2 + ... + \theta_q B^q$. Les paramètres à estimer sont : $\phi_1,...,\phi_p$, $\theta_1,...,\theta_Q$ et σ^2. Pour obtenir des informations sur les méthodes d'estimation de ces paramètres (cf [4]).

3.4.2 Vérification à posteriori et choix du modèle

Vérification à posteriori

- **Tests sur les paramètres :** Dans cette partie, on effectue généralement le test $H_O : \phi_p = 0$ contre l'hypothèse alternative $H_1 : \phi_p \neq 0$, et le test $H_0 : \theta_q = 0$ contre l'hypothèse alternative $H_1 : \theta_q \neq 0$.

  Si $\hat{\phi}_p$ (ou $\hat{\theta}_q$) n'est pas significatif, on relance l'estimation en remplaçant p par $p-1$ (ou q par $q-1$). Pour plus de précisions sur ces tests (cf [4]).

- **Test sur les résidus :** Les résidus estimés (à savoir $\hat{\varepsilon}_t$) sont-ils compatibles avec l'hypothèse de bruit blanc de ε_t ? pour cela, on effectue le test du Portmanteau proposé par Box-Pierce dont la statistique a été améliorée par Ljung-Box. Ce test s'écrit : $H_0 : \hat{\varepsilon}_t$ est un bruit blanc contre l'hypothèse alternative $H_1 : \hat{\varepsilon}_t$ n'est pas un bruit blanc. Pour plus d'informations sur ce test (cf [4]).

Choix du modèle

A l'issue des phases d'estimation et de vérification, il reste en général plusieurs modèles possibles pour représenter les données. Nous choisirons dans cette étude le modèle pour lequel la variance des résidus σ^2 est la plus petite. On montre que l'erreur de prévision diminue avec cette variance. (Cf [1]). Nous fondons notre choix sur ce critère pour la simple raison que l'objectif principal visé dans cette étude est de déterminer un modèle permettant de faire de « bonnes » prévisions.

3.5 Un modèle de simulation en hydrologie : modèle markovien du 1er ordre

Pour faire des simulations, nous utiliserons le modèle markovien du premier ordre décrit dans [8]. Ce type de modèle est approprié pour la simulation des volumes de l'écoulement journalier, mensuel ou annuel. Il s'appuie sur une structure de Markov du premier ordre. C'est-à-dire l'écoulement pour tout intervalle de temps est déterminé par celui de l'intervalle précédent et par une impulsion aléatoire. Un modèle de simulation de ce type appliqué à un processus d'écoulement dont la fréquence des observations est journalière et la périodicité est annuelle, peut s'exprimer comme suit :

$$(*) \qquad X_i = \overline{X_j} + \rho_j \frac{\sigma_j}{\sigma_{j-1}}(X_{i-1} - \overline{X}_{j-1}) + \varepsilon_i \sigma_j \sqrt{1-\rho_j^2}\ ,$$

dans laquelle : $i = 1,2,...,n$; $j = 1,2,...,365$, n est le nombre d'éléments de la simulation que l'on désire effectuer ; X_i est l'écoulement du terme numéro i de la chronique à simuler dont les termes sont numérotés consécutivement à partir de 1 quels que soient le jour de l'année ; j est le numéro du jour de l'année correspondant à la valeur numéro i à simuler ; $\overline{X_j}$ est l'écoulement moyen pour le jour numéro j de l'année ; σ_j est l'écart type des écoulements du jour numéro j de l'année ; ρ_j est le coefficient de corrélation entre les écoulements du jour numéro j et ceux du jour numéro $j-1$ de l'année ; les ε_i sont les réalisations indépendantes d'une variable aléatoire qui suit une loi normale centrée et réduite.

Vu l'équation de simulation $(*)$, il suffit de choisir une valeur initiale de l'écoulement pour simuler une chronique de longueur voulue. Nous utiliserons la chronique de nos données, c'est-à-dire la chronique des débits naturels journaliers dont nous disposons pour estimer les paramètres $\overline{X}_j$, ρ_j, σ_j de ce modèle. Dans les formules ci-dessus, chacun des termes $x_{k,l}$, $k = 1,2,...,365$; $l = 1,2,...,17$, représente le débit du jour k de l'année l. Nous rappelons que notre chronique s'étend sur 17 années.

$$\hat{\bar{X}}_j = \frac{1}{17}\sum_{l=1}^{17} x_{j,l}\ , \quad j = 1,2,3,...,365\ ;$$

$$\hat{\sigma}_j^2 = \frac{1}{17}\sum_{l=1}^{17} (x_{j,l} - \hat{\bar{X}}_j)^2\ , \quad j = 1,2,3,...,365\ ;$$

$$\hat{\rho}_1 = \frac{\sum_{l=1}^{17} (x_{1,l} - \hat{\bar{X}}_1)(x_{365,l} - \hat{\bar{X}}_{365})}{\sqrt{\sum_{l=1}^{17} (x_{1,l} - \hat{\bar{X}}_1)^2}\sqrt{\sum_{l=1}^{17} (x_{365,l} - \hat{\bar{X}}_{365})^2}}\ ;$$

$$\hat{\rho}_j = \frac{\sum_{l=1}^{17} (x_{j,l} - \hat{\bar{X}}_j)(x_{j-1,l} - \hat{\bar{X}}_{j-1})}{\sqrt{\sum_{l=1}^{17} (x_{j,l} - \hat{\bar{X}}_j)^2}\sqrt{\sum_{l=1}^{17} (x_{j-1,l} - \hat{\bar{X}}_{j-1})^2}}\ , \quad j = 1,2,3,...,365\ .$$

Nous avons jugé ce modèle markovien du premier ordre approprié pour faire les simulations de nos débits naturels journaliers pour la simple raison qu'il prend en compte la forte corrélation que nous avons constatée existante entre les débits du jour numéro j et ceux du jour numéro $j-1$ de l'année.

3.6 Evaluation d'un modèle

3.6.1 Erreur d'ajustement et de prévision

Notons $x_{j,c}$, le débit réel du jour j de l'année c ; $\hat{x}_{j,a}$, le débit estimé du jour j de l'année a ; $\hat{x}^*_{j,b}$, le débit prédit du jour j de l'année b.

On rappelle qu'un résidu est la différence entre une valeur estimée ou prédite et sa valeur réelle. Pour bien analyser les résidus, il est donc nécessaire de distinguer l'étude des résidus de l'estimation et celle des résidus de la prévision. En effet, il est évident que les premiers, qui ne sont autres que les erreurs d'ajustement du modèle aux données réelles, sont moins importants et mieux distribués que les seconds, souvent dégradés, qui expliquent pour leur part les erreurs d'ajustement du modèle aux données à prédire.

Pour évaluer la qualité de l'ajustement du modèle, nous calculerons deux types d'erreur globale.

- L'erreur quadratique relative moyenne (er), que nous écrivons pour l'estimation

$$er^2 = \frac{\sum_{j,a}\left[\hat{x}_{j,a} - x_{j,a}\right]^2}{\sum_{j,a} x_{j,a}^2},$$

et pour la prévision de l'année b,

$$er^2 = \frac{\sum_{j,b}\left[\hat{x}^*_{j,b} - x_{j,b}\right]^2}{\sum_{j,b} x^2_{j,b}}.$$

- L'erreur quadratique moyenne (em), que nous écrirons pour l'estimation :

$$em^2 = \frac{\sum_{j,a}\left[\hat{x}_{j,a} - x_{j,a}\right]^2}{n},$$

n est la taille de la série utilisée pour la modélisation ;

et pour la prévision de l'année b,

$$em^2 = \frac{\sum_{j,b}\left[\hat{x}^*_{j,b} - x_{j,b}\right]^2}{n^*}.$$

n^* est le nombre de jours à prédire.

3.6.2 Une méthodologie d'évaluation à posteriori d'un modèle

Nous nous sommes inspirés de la méthodologie proposée dans [6] pour évaluer la fiabilité d'un modèle en termes d'erreurs d'ajustement et de prévision. On suppose qu'un modèle a été préalablement déterminé sur la chronique des débits naturels journaliers de la Sanaga à la station hydrologique de Songmbengué. On désire avoir une idée sur l'évolution annuelle des erreurs quadratiques de prévision et d'ajustement de ce modèle aux débits réels. Pour y parvenir, nous procèderons ainsi :

1- Nous utiliserons premièrement la chronique des débits allant du 1er Juillet 1988 au 30 Juin 2001 pour ré estimer les paramètres de ce modèle, que nous utiliserons pour faire les prévisions de la période allant du 1er Juillet 2001 au 30 Juin 2002. Nous calculerons l'erreur quadratique relative moyenne et l'erreur quadratique moyenne de l'ajustement de ce modèle ré estimé à cette chronique. Puis, nous calculons ces mêmes erreurs pour la prévision effectuée. A l'issue de cette première étape, nous obtenons ainsi les erreurs quadratiques d'adéquation du modèle aux données réelles relatives à l'année 2001-2002.

2- Nous utiliserons deuxièmement la chronique des débits allant du 1er Juillet 1988 au 30 Juin 2002 pour ré estimer les paramètres de ce modèle que nous utiliserons pour faire les prévisions de la période allant du 1er Juillet 2002 au 30 Juin 2003. Nous calculerons l'erreur quadratique relative moyenne et l'erreur quadratique moyenne de l'ajustement de ce modèle

ré estimé à cette chronique. Puis, nous calculons ces mêmes erreurs pour la prévision effectuée. A l'issue de cette deuxième étape, nous obtenons ainsi les erreurs d'adéquation du modèle aux données réelles relatives à l'année 2002-2003.

3- Nous utiliserons troisièmement la chronique des débits allant du 1er Juillet 1988 au 30 Juin 2003 pour ré estimer les paramètres de ce modèle que nous utiliserons pour faire les prévisions de la période allant du 1er Juillet 2003 au 30 Juin 2004. Nous calculerons l'erreur quadratique relative moyenne et l'erreur quadratique moyenne de l'ajustement de ce modèle ré estimé à cette chronique. Puis, nous calculons ces mêmes erreurs pour la prévision effectuée. A l'issue de cette troisième étape, nous obtenons ainsi les erreurs d'adéquation du modèle aux données réelles relatives à l'année 2003-2004.

4- Nous utiliserons quatrièmement la chronique des débits allant du 1er Juillet 1988 au 30 Juin 2004 pour ré estimer les paramètres de ce modèle que nous utiliserons pour faire les prévisions de la période allant du 1er Juillet 2004 au 30 Juin 2005. Nous calculons l'erreur quadratique relative moyenne et l'erreur quadratique moyenne de l'ajustement de ce modèle ré estimé à cette chronique. Puis, nous calculons ces mêmes erreurs pour la prévision effectuée. A l'issue de cette quatrième étape, nous obtenons ainsi les erreurs d'adéquation du modèle aux données réelles relatives à l'année 2004-2005.

CHAPITRE 4

APPLICATIONS ET RESULTATS

4.1 Modélisation et prévision des débits naturels journaliers par un processus ARIMA

Nous adopterons la démarche proposée dans [2] par Box et Jenkins pour modéliser une série chronologique présentant de fortes variations saisonnières comme la nôtre (figure 2.1). Nous présentons cette démarche en l'appliquant à la chronique de nos débits, qui n'est en fait qu'une série chronologique s'étendant sur 17 années, dans laquelle la fréquence des observations est journalière et la périodicité est annuelle. On supposera dans toute cette étude que cette période est de 365 jours. Cette démarche consiste à :

1- Premièrement, éliminer les variations saisonnières de la chronique par standardisation en se servant de la transformation

$$Z_{\nu,\tau} = \frac{X_{\nu,\tau} - \mu_\tau}{\sigma_\tau},$$

où $\nu = 1,2,...,17$; $\tau = 1,2,...,365$; ν est le numéro d'une année ; τ est le numéro d'un jour dans une année ; $X_{\nu,\tau}$ est le débit du jour numéro τ de l'année numéro ν ; μ_τ est la moyenne empirique des débits du jour numéro τ sur les 17 années d'observations, c'est-à-dire

$$\mu_\tau = \frac{1}{17}\sum_{\nu=1}^{17} X_{\nu,\tau} \quad ;$$

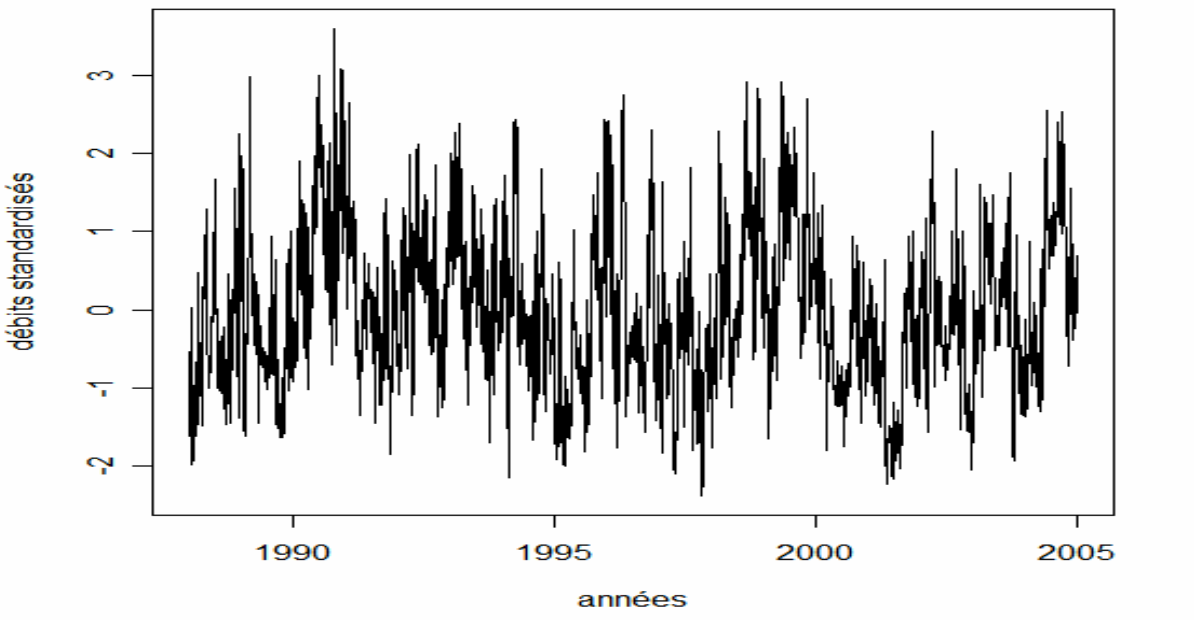
σ_τ^2 est la variance empirique des débits du jour numéro τ sur les 17 années d'observations, c'est-à-dire

$$\sigma_\tau^2 = \frac{1}{17}\sum_{\nu=1}^{17}\left(X_{\nu,\tau} - \mu_\tau\right)^2 .$$

2- Deuxièmement, modéliser la série standardisée $Z_{\nu,\tau}$ par un processus stochastique de type $ARIMA(p,d,q)$.

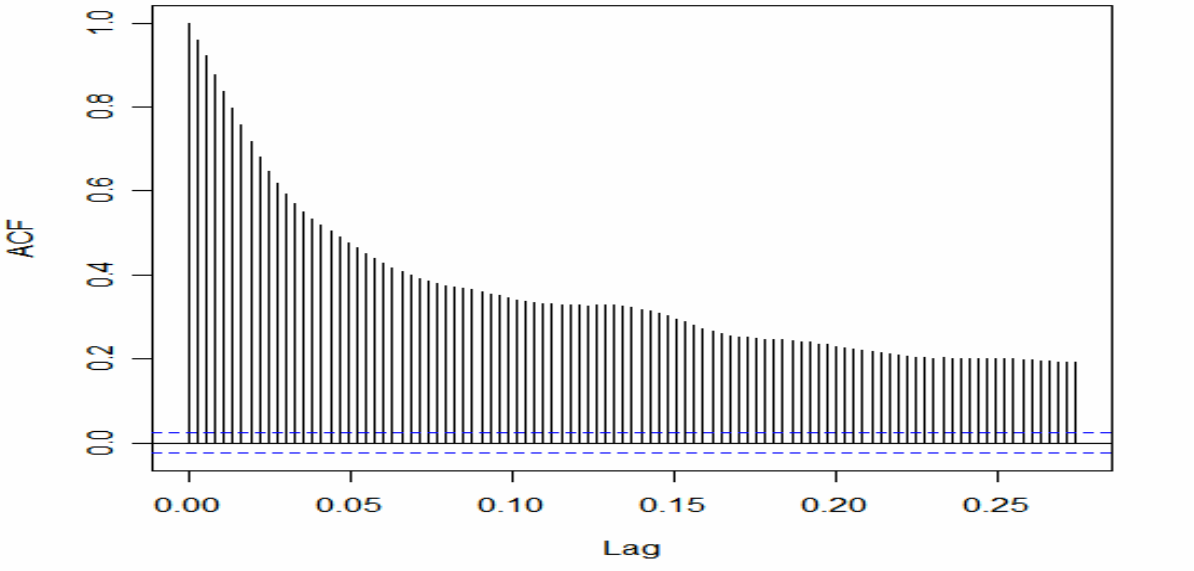
La figure 4.1 présente l'évolution des débits naturels standardisée $Z_{\nu,\tau}$. (Voir commandes C.5 en annexe pour son obtention).

FIG. 4.1 Evolution des débits standardisés



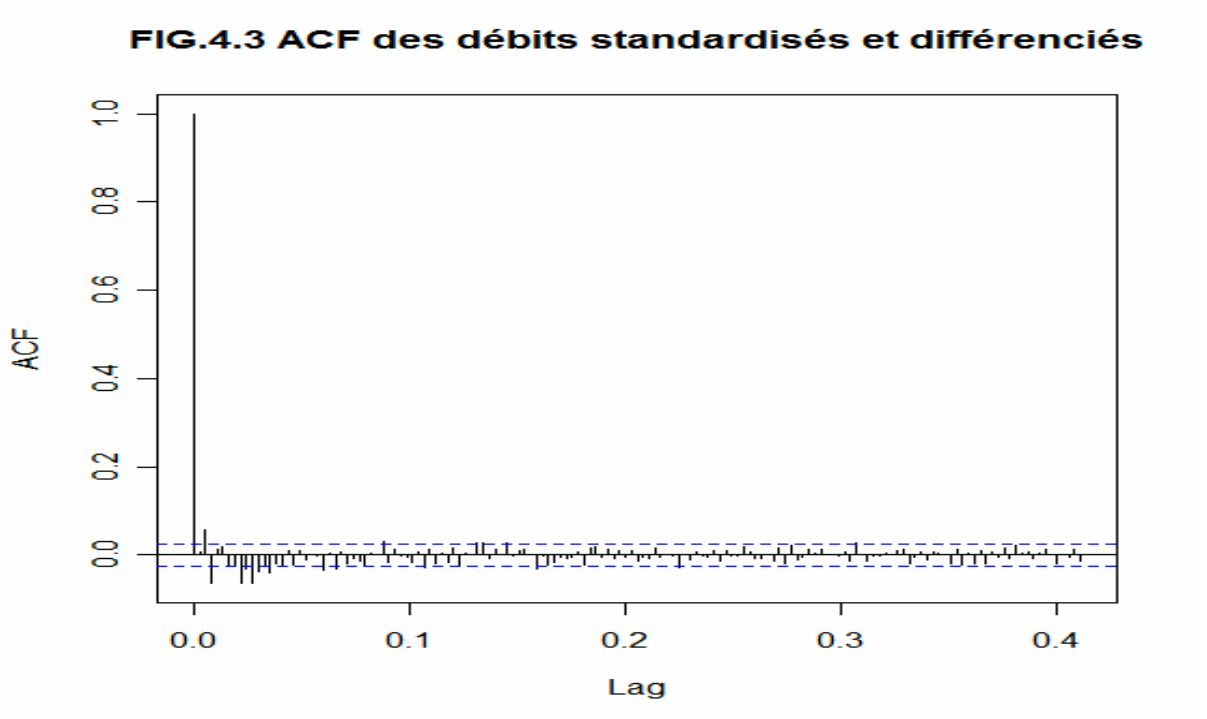
Pour la modélisation du processus $Z_{\nu,\tau}$, nous suivrons la démarche présentée dans la section 3.4 du chapitre 3. En vu d'identifier l'ordre de différentiation d, observons l'auto-corrélogramme (graphe de l'ACF) de la série $Z_{\nu,\tau}$ (figure 4.2). (Voir commande C.6 en annexe pour son obtention).

FIG.4.2 Corrélogramme des débits standardisés



Ce corrélogramme montre une décroissance lente de la fonction d'auto corrélation. On peut donc penser que le processus $Z_t = Z_{365(\nu-1)+\tau} \equiv Z_{\nu,\tau}$ n'est pas stationnaire.

Observons alors l'auto-corrélogramme du processus $Y_t = (1-B)Z_t$ sur la figure 4.3. (Voir commande C.7 en annexe pour son obtention).



Ce corrélogramme montre que la fonction d'autocorrélation décline rapidement vers 0. On peut donc penser que le processus Y_t est stationnaire (une décroissance rapide vers 0 de la fonction d'auto corrélation est une méthode pratique d'identification des processus stationnaires, confère [4]). La probabilité critique du test de non stationnarité de Phillips-Perron est 0.01. Cette valeur étant inférieure à 0.05, on rejette alors l'hypothèse de non-stationarité du processus Y_t avec un risque de se tromper de 5%. (Voir commande C.8 en annexe pour la réalisation de ce test). Nous retenons alors $d = 1$ comme l'ordre de différentiation du processus Z_t.

Rappelons ici quelques résultats pratiques (confère [4]), nous permettant d'apporter une justification supplémentaire pour le choix du nombre 1 comme ordre de différentiation du processus Z_t.

1- Les observations d'un processus stationnaire fluctuent autour d'une valeur moyenne.

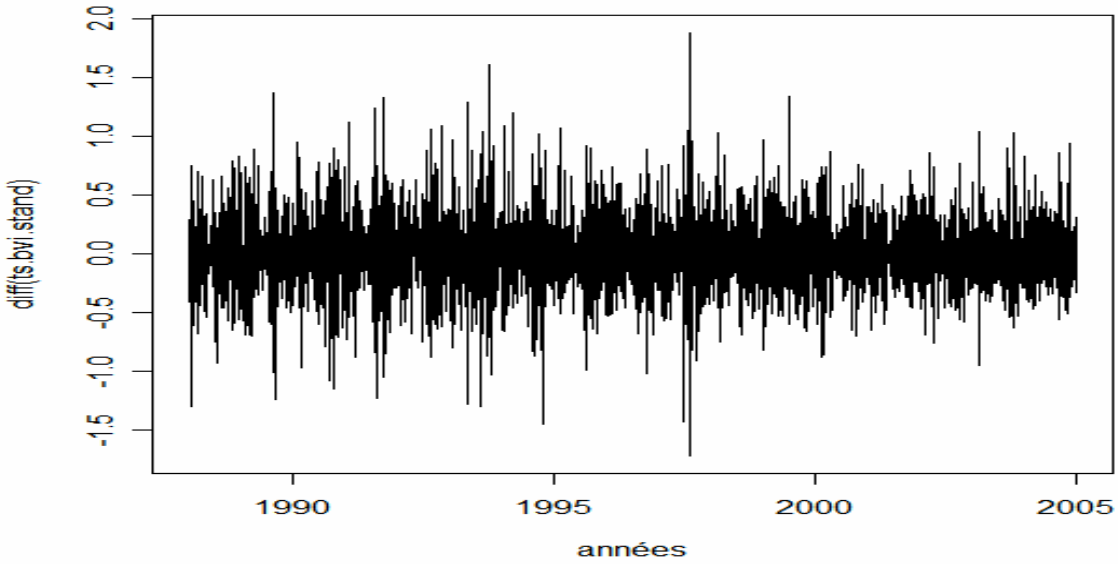
2- Si l'auto corrélation de décalage 1 est égale à 0 ou est négative, le processus n'a pas besoin d'être différencié. Si l'autocorrélation de décalage 1 est inférieure à -0.5, le processus est sur-différencié.

Vérification graphique de ces deux résultats.

On voit clairement sur la figure 4.3 que l'autocorrélation de décalage 1 n'est pas significative. En effet, cette valeur est à l'intérieur de la région de confiance au niveau 95% délimitée par la bande en pointillé bleue.

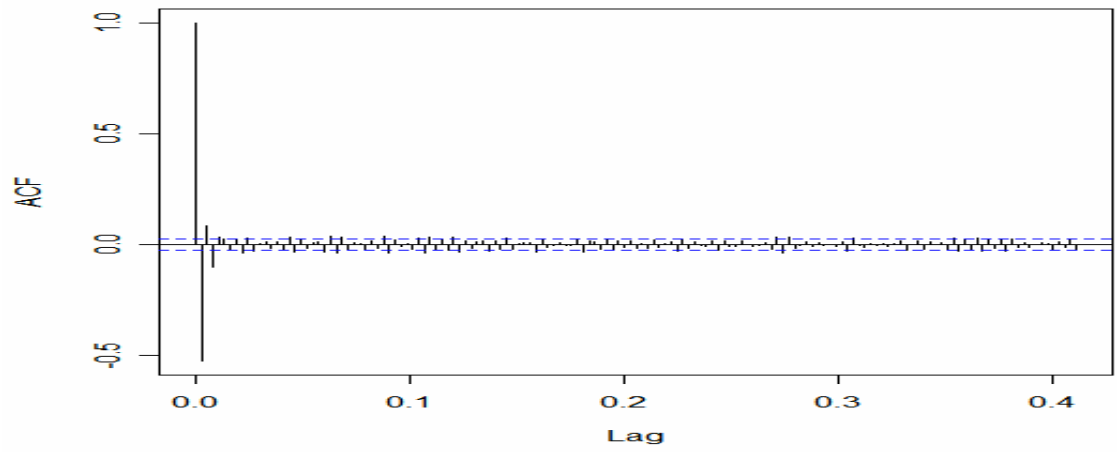
La courbe montrant l'évolution du processus Y_t est présentée sur la figure 4.4 ci-dessous, (voir commande C.9 en annexe pour son obtention). On voit clairement sur cette figure que les observations de ce processus fluctuent autour de la valeur moyenne 0.

FIG. 4.4 Evolution des débits standardisés et différenciés



Considérons le processus $X_t = (1-B)^2 Z_t = (1-B)Y_t$. Son autocorrélogramme est présenté sur la figure 4.5 ci-dessous, (voir commande C.10 en annexe pour son obtention). On voit clairement sur cette figure que l'autocorrélation de décalage 1 est inférieure à -0.5. Ce qui traduit une sur-différentiation du processus Z_t.

FIG.4.5 ACF des débits standardisés et différenciés 2 fois

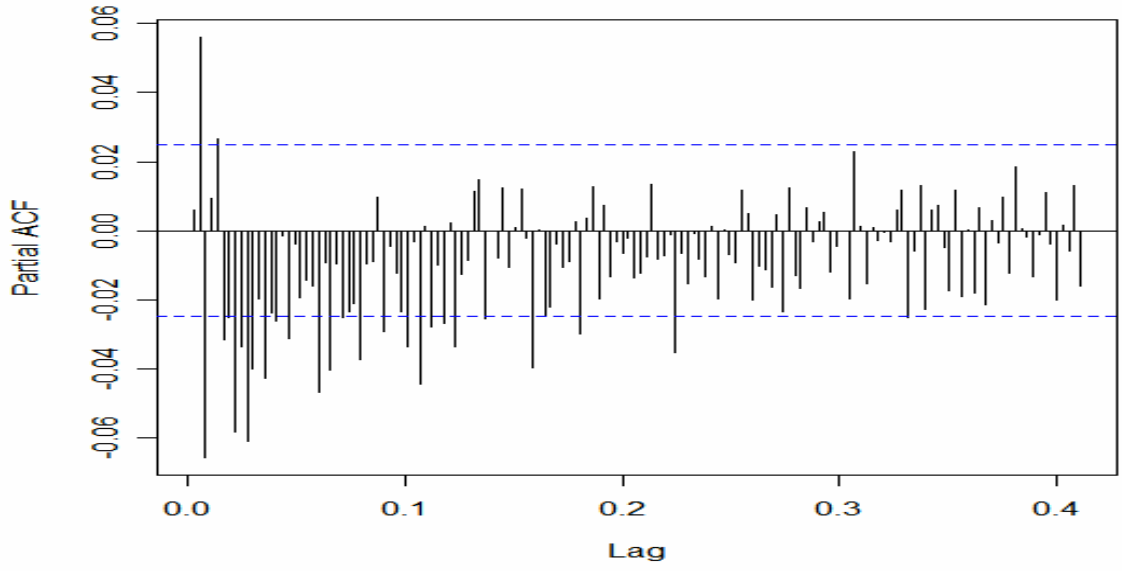


Les résultats précédents confirment le choix de la valeur 1 comme ordre de différentiation du processus Z_t. En vu de déterminer les ordres p et q, nous étudions le processus Y_t tout en supposant qu'il est un $ARMA(p,q)$. Pour cela, nous procédons par une approche empirique qui consiste à déterminer un couple (P,Q) tel que Y_t est approximativement un $AR(P)$ et un $MA(Q)$. Puis, à essayer de poursuivre la modélisation avec tous les couples (p,q) majorés par (P,Q).

Nous choisissons $Q = 13$. En effet, c'est le rang (compté à partir de 0) après lequel les autocorrélations estimées du processus Y_t sont non significatives au seuil de 5% (sur la figure 4.3, les autocorrélations estimées dont les rangs sont supérieurs à 13 sont quasiment toutes à l'intérieur de la bande en pointillé bleue). Cette bande définit une région de confiance au niveau 95%, commune à tous les estimateurs des auto corrélations.

Pour le choix de P, observons la courbe de l'autocorrélogramme partielle (graphe de la fonction d'autocorrélation partielle) du processus Y_t (figure 4.6). (Voir commande C.11 en annexe pour son obtention).

FIG.4.6 PACF des débits standardisés et différenciés



Nous choisissons $P = 75$. En effet, c'est le rang (compté à partir de 1) au-delà duquel les autocorrélations partielles estimées du processus Y_t sont non significatives au seuil de 5% (sur la figure 4.6, les autocorrélations partielles estimées dont les rangs sont supérieurs à 75 sont quasiment toutes à l'intérieur de la bande en pointillé bleue). Cette bande définit une

région de confiance au niveau 95%, commune à tous les estimateurs des auto corrélations partielles.

Remarque 4.1 Notons qu'il existe cependant d'autres méthodes permettant d'estimer conjointement les valeurs de p et q (Confère [1] et [4]) : une première méthode dite *' méthode du coin'* et une seconde méthode dite *'méthode SCAN'*. Ces méthodes ne sont pas, ou du moins, à notre connaissance implémentées dans le logiciel R . La deuxième méthode est implémentée dans le logiciel SAS, mais ce dernier est payant. Le temps imparti à notre étude ne nous a pas permis d'implémenter la première méthode dans R.

Après de multiples modélisations, nous avons constaté que le couple $(14,1)$ fait partie des couples (p,q) majorés par $(75,13)$ ayant rendu possible l'estimation des paramètres du modèle $ARIMA(p,1,q)$ par le logiciel R. De plus, la variance estimée des résidus pour le processus $ARIMA(14,1,1)$ est la plus petite. C'est en fait le critère que nous nous sommes fixés dans la section 3.4.2 du chapitre 3 pour le choix d'un modèle. Nous poursuivons alors cette modélisation en supposant que $p=14$, $q=1$ et $d=1$.

Le modèle $ARIMA(14,1,1)$, défini par le processus Z_t a pour équation :

$$(**) \qquad (1-B)\left(1-\sum_{i=1}^{14} ar_i B^i\right) Z_t = (1+ma_1 B)\varepsilon_t ,$$

où ε_t est un bruit blanc de variance σ^2 ; B^p est l'opérateur retard d'ordre p. Les valeurs estimées des paramètres ar_i, ma_1 et σ^2 , ainsi que leurs intervalles de confiances respectifs sont stockés dans le tableau.4.1 de la page suivante. Ce tableau est obtenu en utilisant les commandes C.12 et C13 en annexe. Ce tableau nous montre d'une part que le coefficient ar14 n'est pas significatif et d'autre part que le coefficient ma1 l'est. Ce qui nous suggère de choisir plutôt un modèle $ARIMA(13,1,1)$.

Le tableau 4.2, obtenu en utilisant les commandes C.14 et C.15 dans l'annexe, contient les paramètres ainsi que les intervalles de confiance des coefficients du modèle $ARIMA(13,1,1)$. Ce tableau montre que le coefficient ar13 n'est pas significatif et que le coefficient ma1 l'est. Ce qui nous suggère de prendre un modèle $ARIMA(12,1,1)$.

TAB. 4.1 Paramètres estimés du modèle $ARIMA(14,1,1)$

	2.5 %		97.5 %
ar1	0.9509	0.9770	1.0030
ar2	0.0130	0.0478	0.0825
ar3	-0.1578	-0.1231	-0.0883
ar4	0.0425	0.0774	0.1124
ar5	-0.0189	0.0161	0.0511
ar6	-0.0874	-0.0525	-0.0175
ar7	-0.0337	0.0013	0.0363
ar8	-0.0651	-0.0302	0.0048
ar9	-0.0117	0.0233	0.0583
ar10	-0.0662	-0.0312	0.0037
ar11	-0.0130	0.0219	0.0568
ar12	-0.0164	0.0184	0.0532
ar13	-0.0582	-0.0235	0.0113
ar14	-0.0044	0.0211	0.0465
ma1	-0.9997	-0.992	-0.9842
σ^2		0.07103	

TAB. 4.2 Paramètres estimés du modèle $ARIMA(13,1,1)$

	2.5 %		97.5 %
ar1	0.9493	0.9755	1.0016
ar2	0.0134	0.0482	0.0829
ar3	-0.1574	-0.122	-0.0879
ar4	0.0418	0.0767	0.1116
ar5	-0.0184	0.0166	0.0515
ar6	-0.0881	-0.0531	-0.0182
ar7	-0.0337	0.0013	0.0363
ar8	-0.0662	-0.0313	0.0037
ar9	-0.0113	0.0236	0.0585
ar10	-0.0646	-0.0298	0.0051
ar11	-0.0155	0.0193	0.0540
ar12	-0.0153	0.0194	0.0542
ar13	-0.0294	-0.0039	0.0216
ma1	-0.9988	-0.9905	-0.9822
σ^2		0.07106	

Le tableau 4.3 ci-dessous, obtenu à l'aide des commandes C.16 et C.17 en annexe, contient les paramètres estimés ainsi que les intervalles de confiance des coefficients du modèle $ARIMA(12,1,1)$. Ce tableau montre que le coefficient ar12 est non significatif et que le coefficient ma1 l'est. Ce qui suggère un modèle $ARIMA(11,1,1)$.

TAB. 4.3 Paramètres estimés du modèle $ARIMA(12,1,1)$

	2.5 %		97.5 %
ar1	0.9495	0.9756	1.0018
ar2	0.0134	0.0481	0.0829
ar3	-0.1574	-0.1226	-0.0878
ar4	0.0417	0.0766	0.1115
ar5	-0.0182	0.0167	0.0517
ar6	-0.0881	-0.0531	-0.0182
ar7	-0.0334	0.0015	0.0364
ar8	-0.0663	-0.0314	0.0036
ar9	-0.0116	0.0233	0.0583
ar10	-0.0640	-0.0293	0.0055
ar11	-0.0157	0.0191	0.0538
ar12	-0.0097	0.0158	0.0413
ma1	-0.9987	-0.9908	-0.9829
σ^2		0.07106	

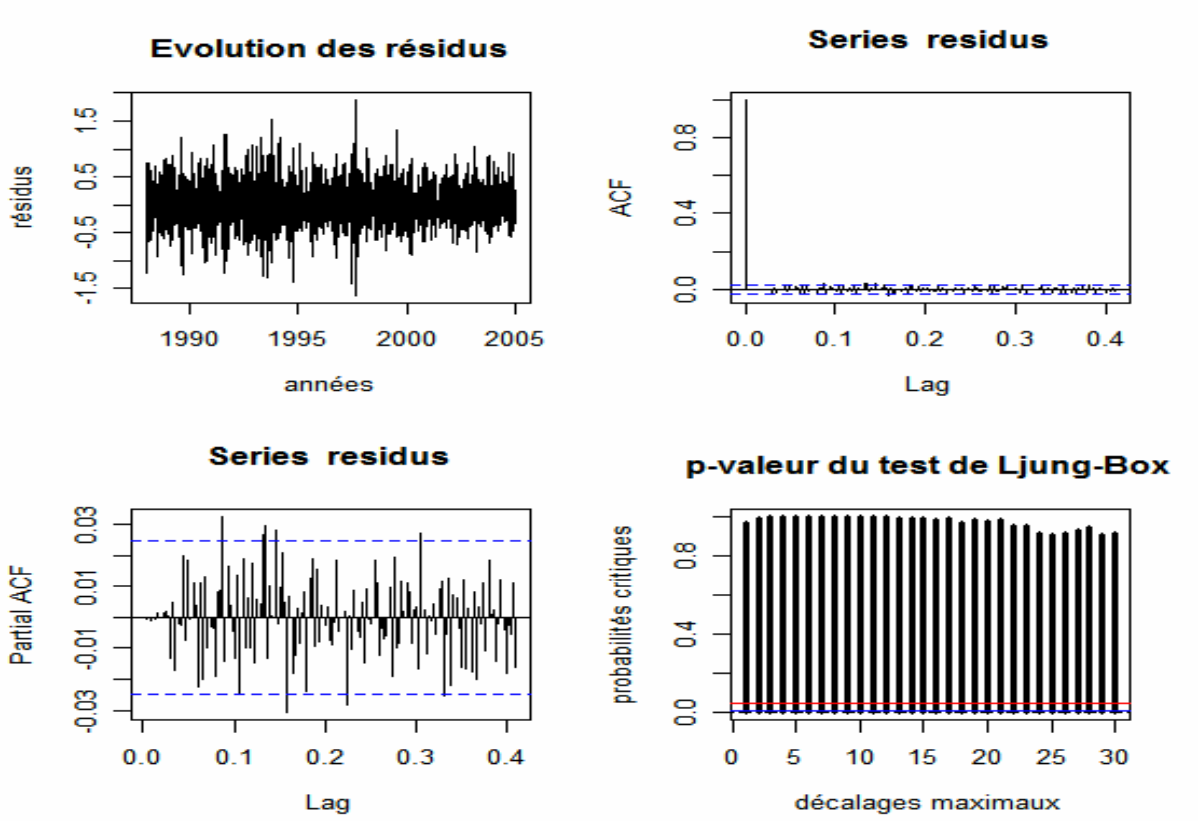
TAB. 4.4 Paramètres estimés du modèle $ARIMA(11,1,1)$

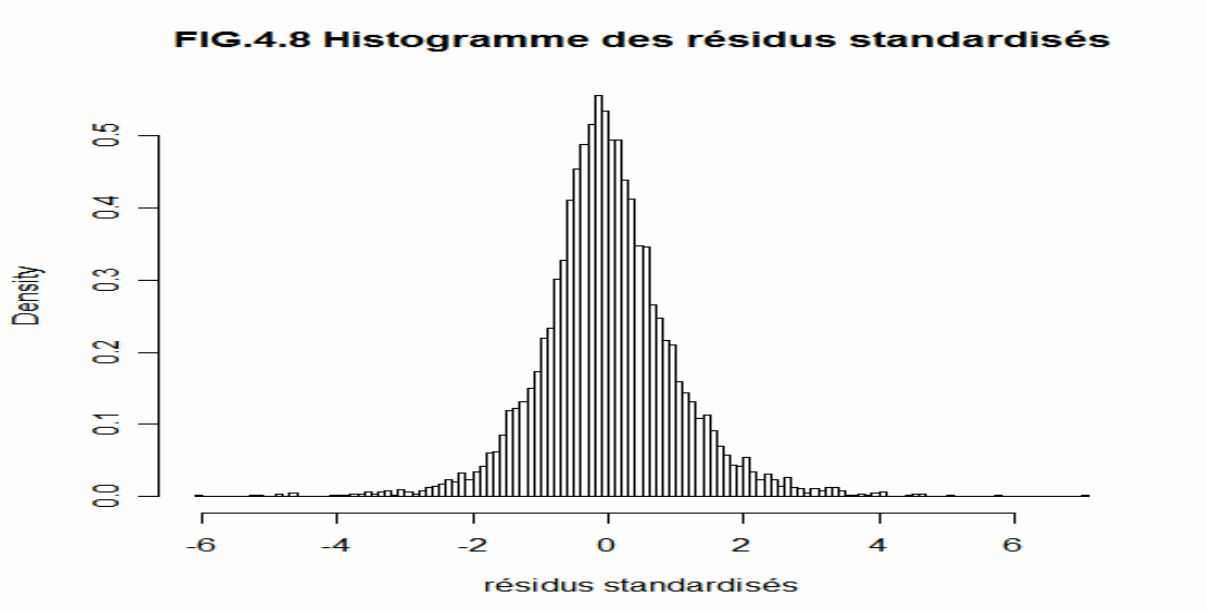
	2.5 %		97.5 %
ar1	0.9488	0.9751	1.0013
ar2	0.0128	0.0476	0.0823
ar3	-0.1569	-0.1222	-0.0874
ar4	0.0412	0.0761	0.1109
ar5	-0.0182	0.0167	0.0516
ar6	-0.0889	-0.0539	-0.0190
ar7	-0.0332	0.0017	0.0366
ar8	-0.0651	-0.0302	0.0047
ar9	-0.0134	0.0214	0.0561
ar10	-0.0632	-0.0285	0.0062
ar11	0.0081	0.0337	0.0593
ma1	-0.9979	-0.9897	-0.9815
σ^2		0.07108	

Le tableau 4.4 ci-dessus, contient les paramètres estimés ainsi que les intervalles de confiance des coefficients du modèle $ARIMA(11,1,1)$ et est obtenu en utilisant les commandes C.18 et C.19 dans l'annexe. Ce tableau montre que les coefficients ar11 et ma1 sont significatifs. Ce qui nous permet de retenir finalement le modèle $ARIMA(11,1,1)$.

Vérifions l'hypothèse de bruit blanc des résidus. La courbe décrivant l'évolution des résidus, l'auto-corréllogramme et l'autocorrélogramme partielle des résidus (voir figures 4.7 obtenue en utilisant la commande C.20 en annexe), permettent de conclure que le processus ε_t est stationnaire. En vue de vérifier l'hypothèse de non corrélation des résidus, nous effectuons le test de Ljung-Box dont l'hypothèse nulle est la nullité de tous les auto-corrélations de décalages inférieurs ou égaux à un certain ordre que l'on précise lors de la réalisation de ce test dans le logiciel R. Ainsi, pour vérifier l'hypothèse d'indépendance entre les résidus, nous construisons un graphe présentant les valeurs des probabilités critiques du test de Ljung-Box en fonction des ordres maximaux de décalages fixés (figures 4.7). On observe sur ce graphe que ces probabilités critiques sont toutes supérieures à 0.05, ce qui nous permet alors de ne pas rejeter l'hypothèse nulle au seuil 5%. Nous pouvons dont conclure que ces résidus forment effectivement un bruit blanc.

FIG. 4.7 Diagnostic des résidus ε_t du modèle $ARIMA(11,1,1)$





Nous présentons sur la figure 4.8 ci-dessus, obtenue en utilisant la commande C.21 en annexe, l'histogramme du processus $\eta_t = \frac{\varepsilon_t}{\sigma}$, où σ est l'écart-type du processus stationnaire ε_t. Cet histogramme présente l'allure d'une distribution normale.

Vu la non significativité des coefficients ar5, ar7, ar8, ar9 et ar10 (tableau 4.4), nous concluons que le processus Z_t suit un modèle $ARIMA(11,1,1)$ dont l'équation est :

$(1-0,9751B-0,0476B^2+0,1222B^3-0,0761B^4+0,0539B^6-0,0337B^{11})Z_t=(1-0,9897B)\varepsilon_t$

où ε_t est un bruit blanc de variance estimée $\hat{\sigma}^2 = 0,07108$.

Nous pouvons à présent estimer le processus Z_t par le modèle $ARIMA(11,1,1)$ et par la suite, à l'aide de la transformation $X_{\nu,\tau} = \sigma_\tau Z_{\nu,\tau} + \mu_\tau$ estimer la chronique de nos débits $X_{\nu,\tau}$.

Conclusion générale de cette modélisation

Nous terminons cette partie en disant que l'équation finale du modèle vérifié par la chronique de nos débits naturels journaliers $X_{\nu,\tau}$ est

$(***)$ $$X_{\nu,\tau} = \sigma_\tau Z_{\nu,\tau} + \mu_\tau,$$

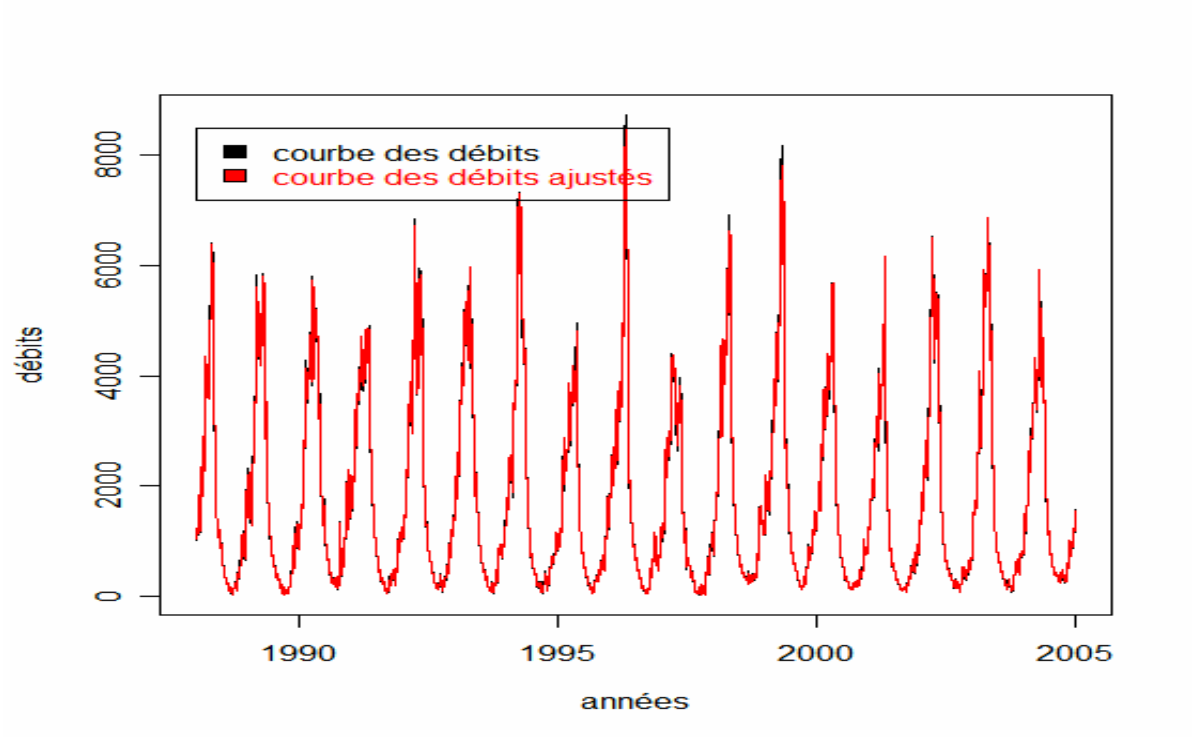
où $Z_{\nu,\tau}$ est un processus qui suit le modèle $ARIMA(11,1,1)$ dont les paramètres sont stockés dans le tableau 4.4.

Nous utiliserons cette équation sous la forme suivante pour estimer les valeurs de la chronique de ces débits

$$\widehat{X}_{\nu,\tau} = \sigma_\tau \widehat{Z}_{\nu,\tau} + \mu_\tau ,$$

où $\nu = 1,2,...,17$, $\tau = 1,2,...,365$; les $\widehat{Z}_{\nu,\tau}$ sont les valeurs du processus $Z_{\nu,\tau}$ estimées par le modèle $ARIMA(11,1,1)$. Nous présentons sur la figure 4.9 obtenue en utilisant la commande C.22 dans l'annexe, l'ajustement de ce modèle final à la chronique des débits réels.

FIG. 4.9 Ajustement de la série des débits $X_{\nu,\tau}$ par le modèle final



Nous présentons dans le tableau 4.5 obtenu en utilisant la commande C.23 dans l'annexe, les erreurs quadratiques liées à cet ajustement. Notons que la chronique de ces débits s'étend sur 6205 jours.

TAB. 4.5 Erreurs quadratiques de l'ajustement du modèle final aux débits $X_{\nu,\tau}$

	Nombre de jours	er (%)	em (m^3/s)
Ajustement	6205	4.91	120.3

Pour faire les prévisions des débits naturels journaliers d'une année k fixée, notée $\widehat{X}^*_{k,\tau}$, nous utilisons la formule suivante

$$\widehat{X}^*_{k,\tau} = \sigma_\tau \widehat{Z}^*_{k,\tau} + \mu_\tau ,$$

où $\tau = 1,2,...,365$; $\hat{Z}^{*}_{k,\tau}$ sont les prévisions de l'année k effectuées avec le modèle $ARIMA(11,1,1)$ dont les coefficients sont stockés dans le tableau 4.4.

Evaluation a posteriori de la fiabilité du modèle

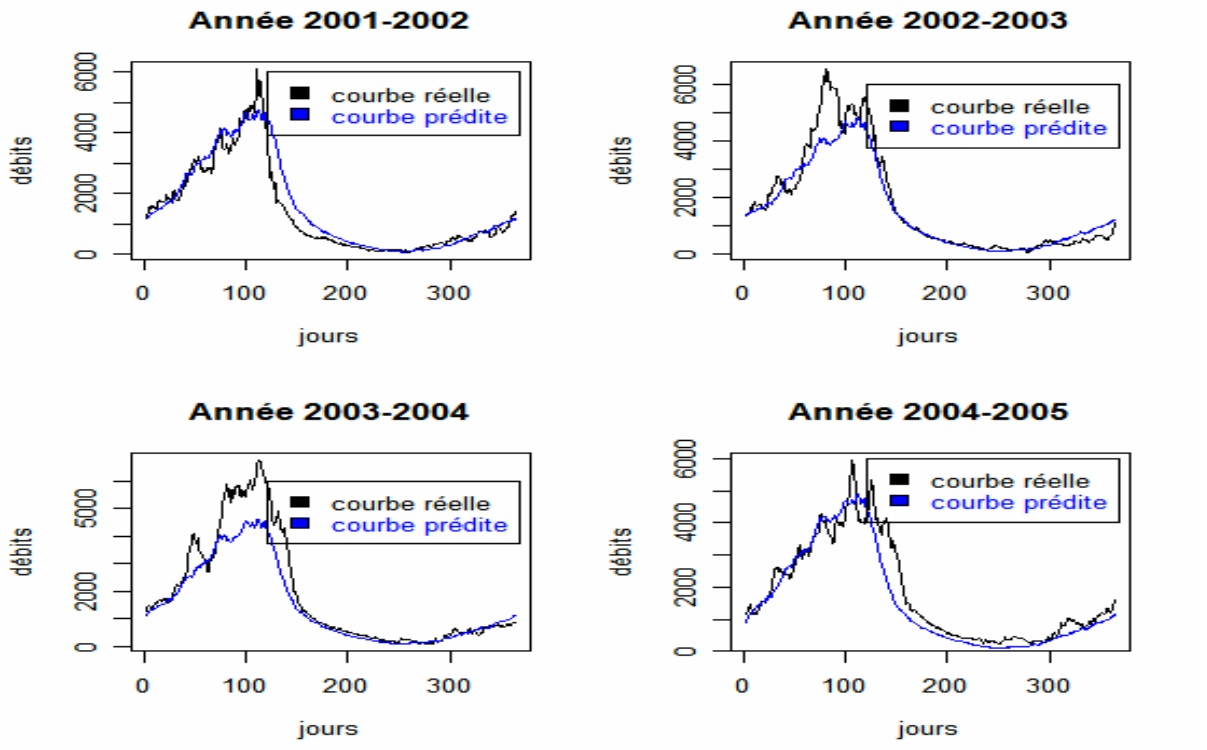
Afin d'évaluer la fiabilité de ce modèle final, nous allons suivre le principe décrit dans la section 3.6.2 du chapitre 3. Les résultats sont stockés dans le tableau 4.6 obtenu en utilisant la commande C.24 en annexe.

TAB. 4.6 Erreurs quadratiques des ajustements et des prévisions a posteriori

	Ajustement	**Ajustement**	**Ajustement**	**Prévision**	**Prévision**	**Prévision**
Années	Nombre de jours	er (%)	em (m^3/s)	Nombre de jours	er (%)	em (m^3/s)
2001-2002	4745	5.5	137.5	365	21.4	433.3
2002-2003	5110	5.4	132.5	365	22.1	555.1
2003-2004	5475	5.2	128.0	365	27.0	730.5
2004-2005	5840	5.0	124.0	365	21.3	493.8

Chacun des graphiques de la figure 4.10 ci-dessous obtenue en utilisant la commande C.25 dans l'annexe, présente la courbe des débits réels sur laquelle est superposée celle prédite par le modèle décrit par l'équation $(***)$ où les paramètres σ_τ, μ_τ et ceux du modèle $ARIMA(11,1,1)$ sont ré estimés suivant le principe décrit à la section 3.6.2 du chapitre 3.

FIG. 4.10 Courbes des débits réels et des prévisions



4.2 Modélisation et prévision des débits naturels journaliers par un modèle multiplicatif

La deuxième approche que nous allons adoptée pour modéliser les débits naturels journaliers du B.V.I. de la Sanaga à la station de contrôle de Songmbengué consiste à caractériser ces derniers par un modèle multiplicatif en partant d'une désaisonnalisation de cette chronique à l'aide de la méthode des moyennes mobiles centrées. Pour y parvenir, nous suivrons la méthode décrite à la section 3.2 du chapitre 3.

La droite de régression des écarts-types en fonction des moyennes des débits de chacune des années étudiées a pour équation

$$\sigma = 0.89\overline{x} + 129.7,$$

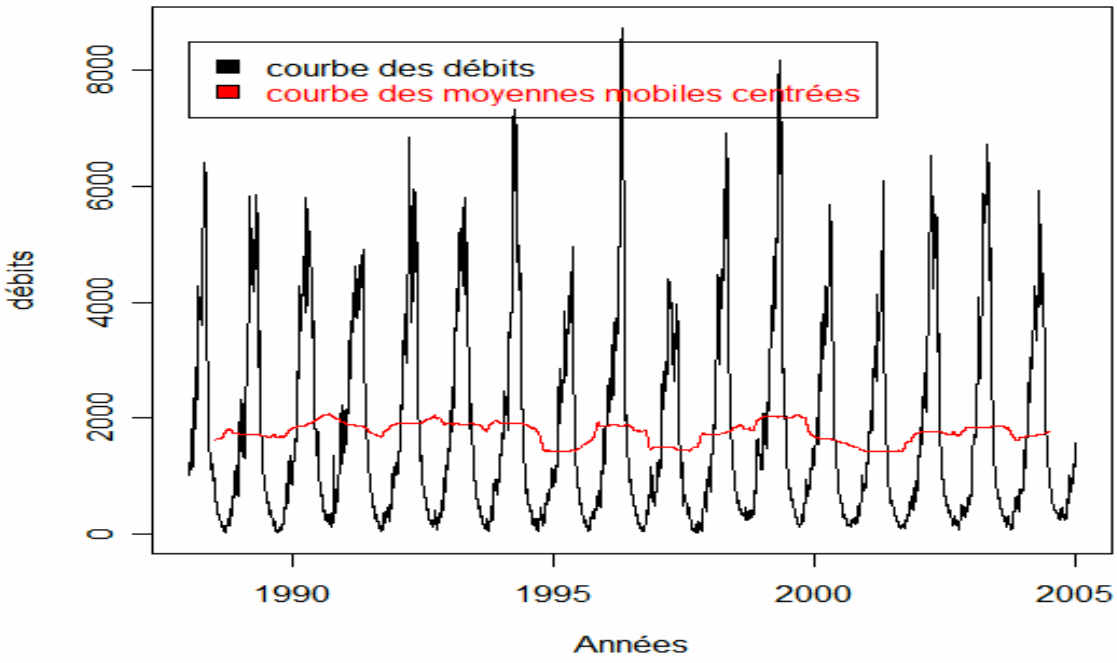
où le coefficient 0.89 est très significatif (sa probabilité critique est 0.000373), (voir commande C.26 en annexe pour l'obtention de ces résultats). Ce qui suggère d'adopter effectivement un schéma de composition multiplicatif pour ces débits naturels journaliers. Rappelons que la forme générale de l'équation de ce type de modèle est :

$$y_t = C_t \times S_t \times \varepsilon_t,$$

où (y_t) est la série des débits que nous étudions, C_t est la composante tendancielle, S_t est la composante saisonnière, ε_t est la composante résiduelle.

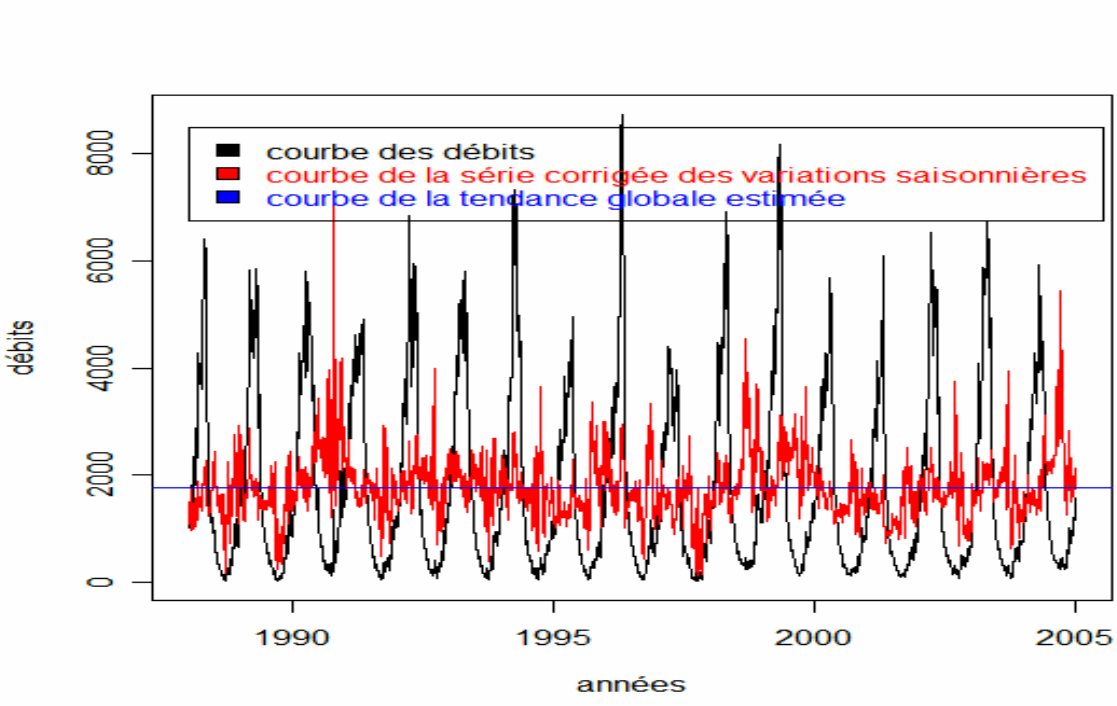
La figure 2.1 du chapitre 2 nous montre clairement que cette chronique des débits est quasiment formée des variations saisonnières. C'est-à-dire sa tendance varie très peu. Ainsi, afin d'expurger cette série de ses variations intra annuelles périodiques, c'est-à-dire de désaisonnaliser cette série, nous utilisons comme technique mathématique la méthode des moyennes mobiles centrées qui à l'avantage de ne faire aucune hypothèse à priori sur la forme de la tendance à estimer. L'ordre de la moyenne mobile que nous utilisons pour désaisonnaliser cette chronique est de 365. En effet, la saisonnalité ou la période de cette série est de 365 jours. La figure 4.11 présente la courbe décrite par la série des moyennes mobiles centrées d'ordre 365, et celle de la série des débits naturels. Cette figure est obtenue en utilisant la commande C.27 dans l'annexe.

FIG.4.11 Série des moyennes mobiles d'ordre 365



Les coefficients saisonniers corrigés sont stockés dans le tableau 4.15 en annexe. Ils sont obtenus en utilisant la commande C.28 dans l'annexe. La figure 4.12 ci-dessous présente la série corrigée des variations saisonnières, qu'une estimation de la tendance globale à long terme de la chronique des débits. Cette figure est obtenue en utilisant la commande C.29 en annxe.

FIG. 4.12 Graphe de la série corrigée des variations saisonnières



On voit clairement sur la figure 4.12 que la tendance globale à long terme de cette chronique est presque constante. Cette tendance a été estimée en effectuant la régression de la série corrigée des variations saisonnières en fonction du temps. L'équation de cette tendance estimée est :

$$C_t = 0,0059t + 1751,$$

où les coefficients 0,0059 et 1751 ont pour probabilités critiques respectives 0,14 et 2×10^{-16}. Ce qui permet de conclure que la pente de cette droite n'est pas significative et que son ordonnée à l'origine est très hautement significative au seuil de 5%. Nous avons utilisé la commande C.30 en annexe pour obtenir ces résultats. L'équation générale de ce modèle est :

$$y_t = (\alpha + \beta.t) \times (\sum_{i=1}^{365} \gamma_i S_t^i) \times \varepsilon_t, \qquad \text{où} \quad S_t^i = \begin{cases} 1, \text{ si } t \text{ correspond au jour i} \\ 0, \text{ sinon} \end{cases}.$$

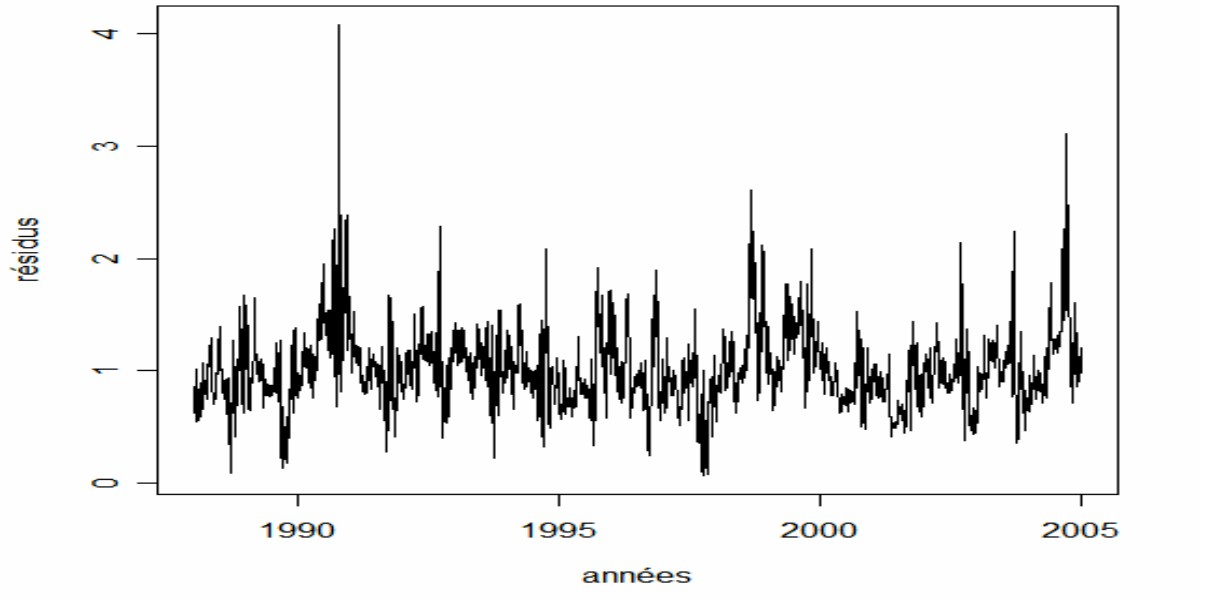
Compte tenu des résultats qui précèdent, ce modèle estimé à alors la forme suivante :

$$\hat{y}_t = 1751 \times \sum_{i=1}^{365} \hat{\gamma}_i S_t^i,$$

où les coefficients $\hat{\gamma}_i$ sont stockés dans le tableau 4.15.

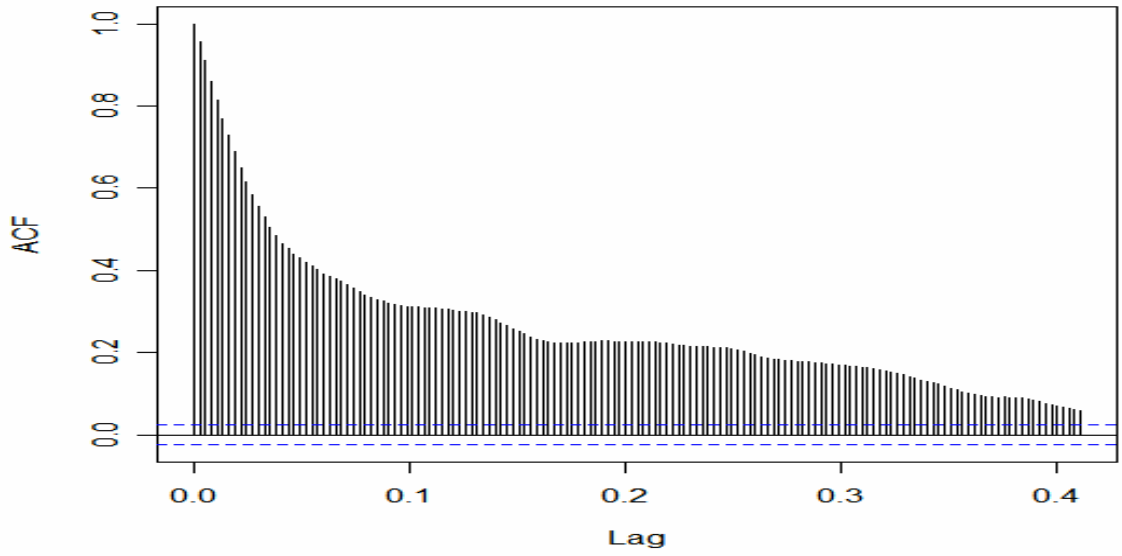
La question que l'on pourrait se poser est celle de savoir si les résidus résultant de l'ajustement des débits par ce modèle estimé forment un bruit blanc ? Pour répondre à cette question, étudions dont le processus ε_t à travers ses valeurs estimées $\hat{\varepsilon}_t = \frac{y_t}{\hat{y}_t}$.

FIG.4.13 Graphe des résidus



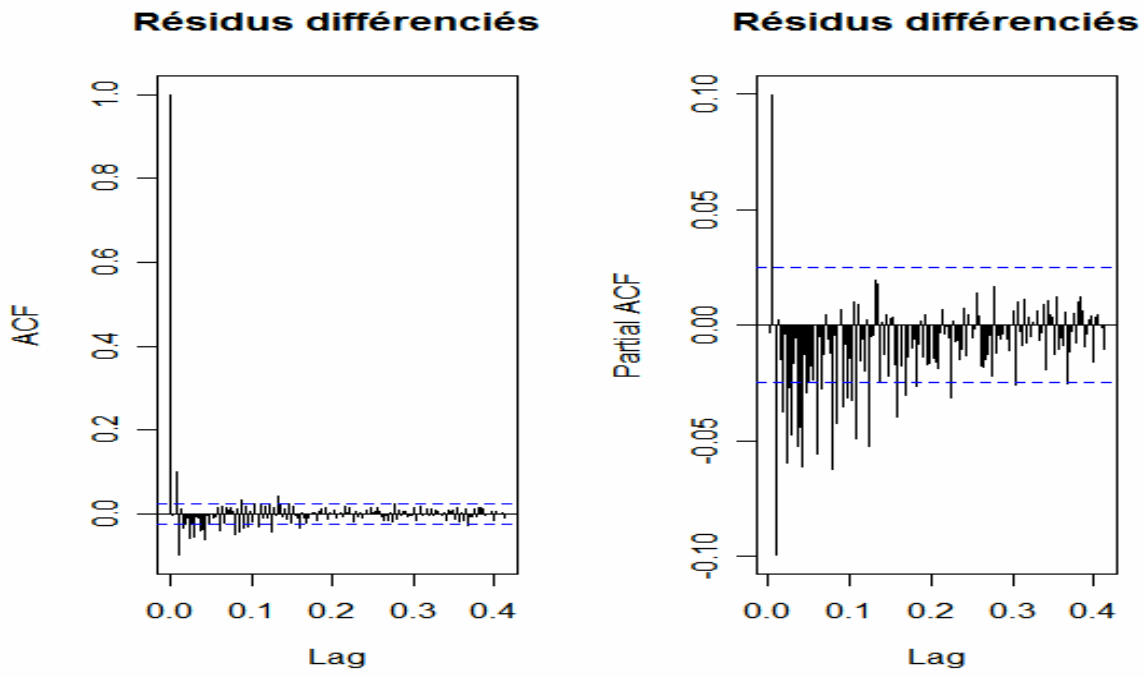
La figure 4.13 ci-dessus présente le graphe des résidus estimés $\hat{\varepsilon}_t$. Elle est obtenue en utilisant la commande C.31 dans l'annexe. Le graphe des autocorrélations présentées sur la figure 4.14 montre que le processus ε_t n'est pas stationnaire (décroissance lente des autocorrélations). ε_t n'est dont pas un bruit blanc Cette figure est obtenue en utilisant la commande C.32 dans l'annexe.

FIG.4.14 Résidus estimés



Nous présentons sur la figure 4.15 le corrélogramme et le corrélogramme partiel du processus $w_t = (1-B)\varepsilon_t$. Ces graphes sont obtenues en utilisant la commande C.33 en annexe.

FIG. 4.15 Corrélogrammes du processus w_t



Le processus w_t paraît stationnaire vu les graphes de la figure 4.15. Nous nous proposons alors de modéliser le processus ε_t par un modèle $ARIMA(p,1,q)$. L'observation des auto-corrélogrammes de figure 4.15 permet de majorer les couples (p,q) possibles par le couple (40,33). Après de multiples modélisations avec différentes valeurs du couple (p,q) vérifiant cette condition, nous avons retenu le couple (16 ,1) comme celui pour lequel la variance des résidus du modèle est minimale. Le modèle $ARIMA(16,1,1)$ vérifié par le processus ε_t a pour équation générale

$$\sum_{i=1}^{16}(1-ar_iB^i)(1-B)\varepsilon_t = (1+ma_1B)\eta_t \ ,$$

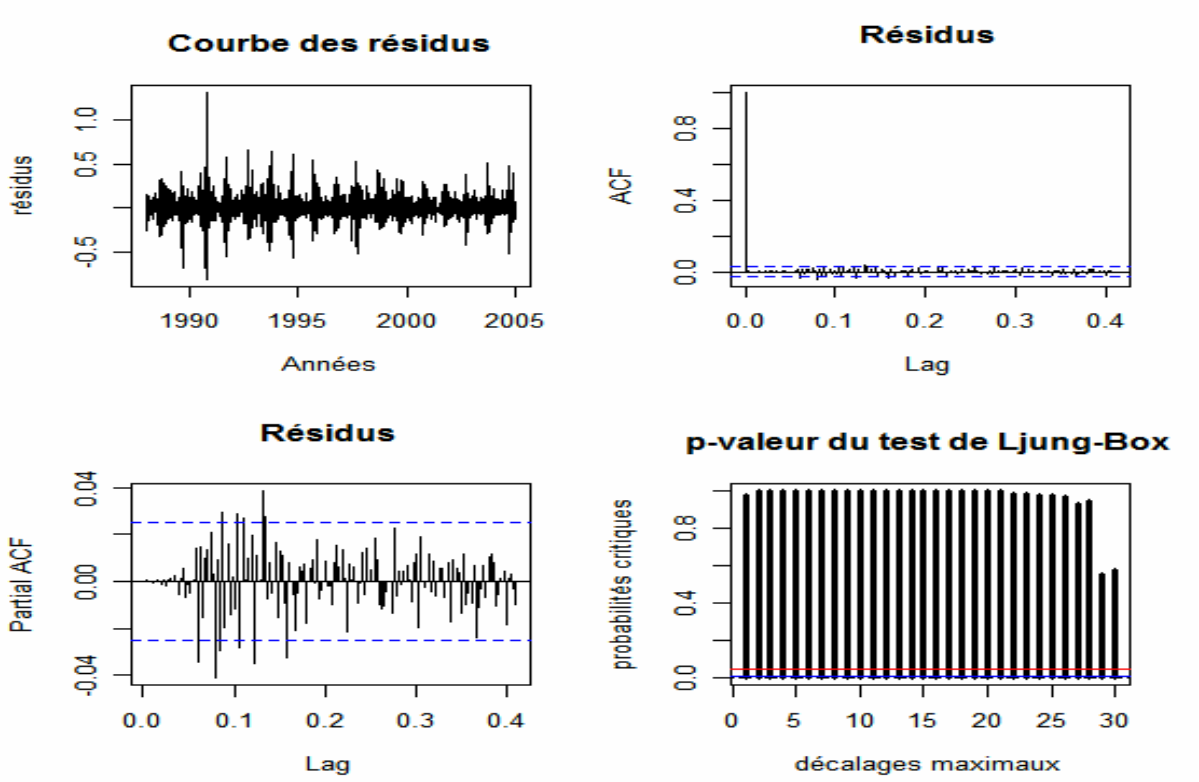
où η_t est un bruit blanc de variance σ_η^2, les paramètres estimés de ce modèle ainsi que les intervalles de confiance correspondants son stockés dans le tableau 4.7 ci-dessous. Ce tableau est obtenu en utilisant les commandes C.34 et C.35 dans l'annexe. On peut noter sur ce tableau que les coefficients ar16 et ma1 sont significatifs. On peut noter sur ce tableau que les coefficients ar16 et ma1 sont significatifs.

TAB. 4.7 Paramètres estimés du modèle $ARIMA(16,1,1)$

	2.5 %		97.5 %
ar1	0.9474	0.9736	0.9997
ar2	0.0548	0.0895	0.1242
ar3	-0.2287	-0.1939	-0.1592
ar4	0.0682	0.1033	0.1384
ar5	-0.0627	-0.0275	0.0076
ar6	-0.0472	-0.012	0.0232
ar7	-0.0846	0.0258	0.0610
ar8	-0.0846	-0.0494	-0.0143
ar9	-0.0064	0.0288	0.0640
ar10	-0.0613	-0.0261	0.0091
ar11	0.0007	0.0358	0.0710
ar12	-0.0270	0.0082	0.0434
ar13	-0.0767	-0.0416	-0.0065
ar14	-0.0315	0.0033	0.0380
ar15	-0.0527	-0.0180	0.0166
ar16	0.0135	0.0391	0.0646
ma1	-0.9999	-0.9918	-0.9838
σ_η^2		0.0089	

Vérifions l'hypothèse de bruit blanc des résidus η_t de ce modèle. On observe sur les graphes de la figure 4.16 obtenue en utilisant la commande C.36 en annexe, que ces résidus forment effectivement un bruit blanc.

FIG. 4.16 Diagnostic des résidus η_t



Compte tenu de la non significativité des coefficients ar5, ar6, ar7, ar9, ar10, ar12, ar14, ar15 (tableau 4.7), le modèle défini par le processus ε_t a finalement pour équation

$$(1-B)(1-0,97356B-0,0895B^2+0,1939B^3-0,1033B^4+0,0494B^8-0,0358B^{11}+0,0416B^{13}-0,0394B^{16})\varepsilon_t=(1-0,9918B)\eta_t$$

où η_t est un bruit blanc de variance $\sigma_\eta^2 = 0.0089$.

Conclusion générale de cette modélisation

Nous terminons cette partie en disant que l'équation finale du modèle vérifié par la chronique de nos débits naturels journaliers y_t est

$$(****) \qquad y_t = (\alpha+\beta t)\times\left(\sum_{i=1}^{365}\gamma_i S_t^i\right)\times\varepsilon_t,$$

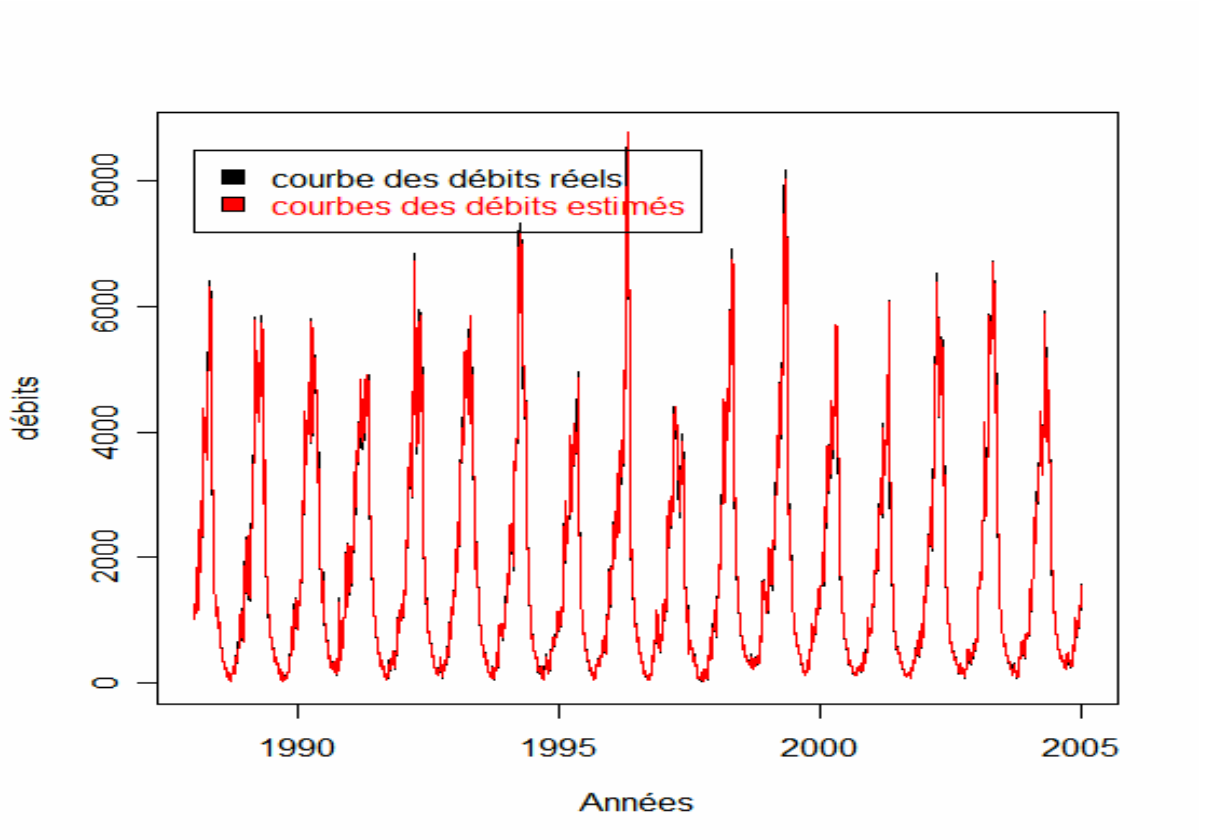
où ε_t est un processus qui suit le modèle $ARIMA(16,1,1)$ dont les paramètres sont stockés dans le tableau 4.7.

Nous utiliserons cette équation sous la forme suivante pour estimer les valeurs de la chronique de ces débits

$$\hat{y}_t = \hat{\alpha} \times \left(\sum_{i=1}^{365} \hat{\gamma}_i S_t^i \right) \times \hat{\varepsilon}_t ,$$

où les $\hat{\varepsilon}_t$ sont les valeurs du processus ε_t estimées par le modèle $ARIMA(16,1,1)$. Nous présentons sur la figure 4.17 obtenue en utilisant la commande C.37 dans l'annexe, l'ajustement de ce modèle final à la chronique des débits réels.

FIG. 4.17 Ajustement des débits y_t



Nous présentons dans le tableau 4.8 obtenu en utilisant la commande C.38 dans l'annexe, les erreurs quadratiques liées à cet ajustement. Notons que la chronique de ces débits s'étend sur 6205 jours.

TAB. 4.8 Erreurs quadratiques de l'ajustement du modèle final aux débits y_t

	Nombre de jours	er (%)	em (m^3/s)
Ajustement	6205	5.1	126.8

Pour faire les prévisions des débits naturels journaliers d'une année k fixée, notée $\hat{y}_t^*$, nous utilisons la formule suivante

$$\hat{y}_t^* = \hat{\alpha} \times \left(\sum_{i=1}^{365} \hat{\gamma}_i S_t^i \right) \times \hat{\varepsilon}_t^* ,$$

où $t = 1, 2, ..., 365$; $\hat{\varepsilon}_t^*$ sont les prévisions des débits de l'année k effectuées avec le modèle $ARIMA(16,1,1)$ dont les coefficients sont stockés dans le tableau 4.7.

Evaluation a posteriori de la fiabilité du modèle

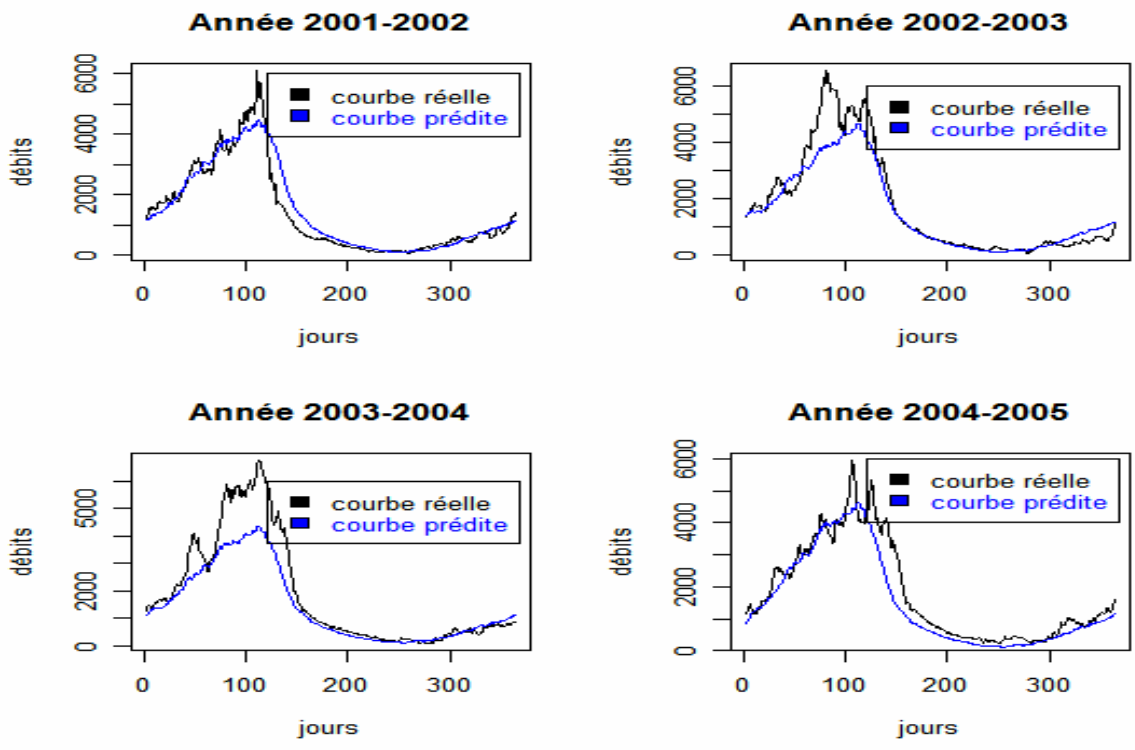
Afin d'évaluer la fiabilité de ce modèle final, nous allons suivre le principe décrit dans la section 3.6.2 du chapitre 3. Les résultats sont stockés dans le tableau 4.9 obtenu en utilisant la commande C.39 dans l'annexe.

TAB. 4.9 Erreurs quadratiques des ajustements et des prévisions a posteriori

	Ajustement	**Ajustement**	**Ajustement**	**Prévision**	**Prévision**	**Prévision**
Années	Nombre de jours	er (%)	em (m^3/s)	Nombre de jours	er (%)	em (m^3/s)
2001-2002	4745	5.2	130.4	365	20.6	417.3
2002-2003	5110	5.3	129.4	365	23.9	600.8
2003-2004	5475	5.2	129.0	365	30.4	823.3
2004-2005	5840	5.2	128.0	365	21.9	506.6

Chacun des graphiques de la figure 4.18 ci-dessous obtenue en utilisant la commande C.40 en annexe, présente la courbe des débits réels sur laquelle est superposée celle prédite par le modèle décrit par l'équation $(****)$, où les paramètres α, β, γ_i et ceux du modèle $ARIMA(16,1,1)$ sont ré estimés suivant le principe décrit à la section 3.6.2 du chapitre 3.

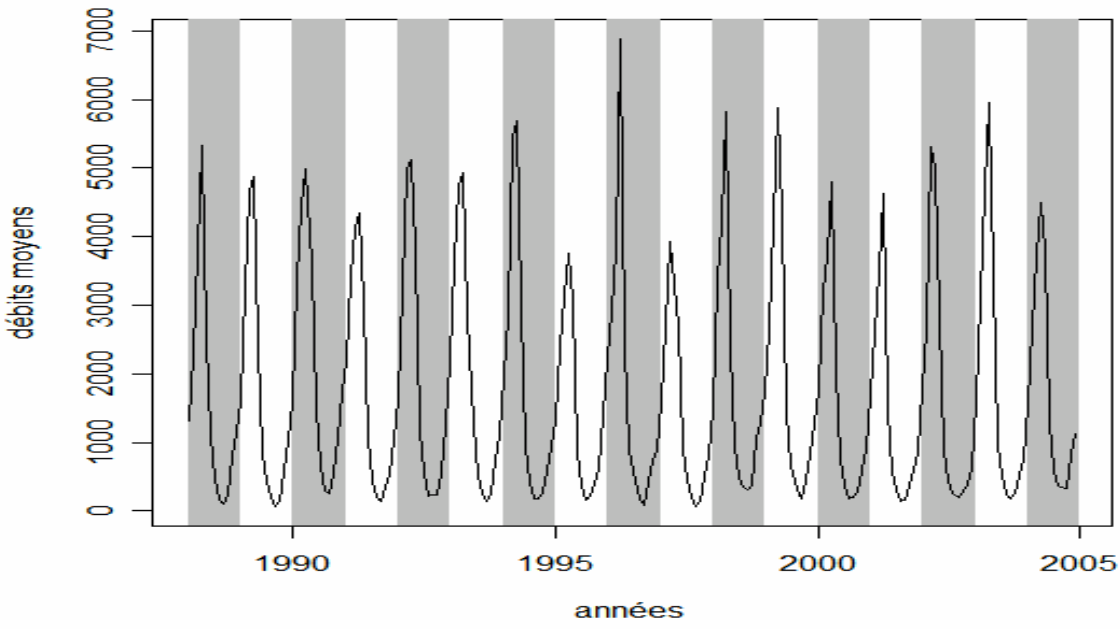
FIG. 4.18 Courbes des débits réels et des prévisions



4.3 Modélisation des débits moyens mensuels du B.V.I. de la Sanaga par un modèle de type SARIMA

Nous modéliserons dans cette partie les débits moyens mensuels correspondants aux débits naturels journaliers du B.V.I. de la Sanaga à la station hydrologique de Songmbengué. Nous présentons sur la figure 4.19 obtenue en utilisant la commande C.41 en annexe, la courbe de ces débits moyens mensuels. On remarque sur cette courbe que la chronique des débits moyens mensuels est aussi dominée par les variations saisonnières. Considérons le processus Y_t dont les observations forment la chronique des débits moyens mensuels. Nous présentons sur la figure 4.20 obtenue en utilisant la commande C.42 en annexe, le corrélogramme et le corrélogramme partiel de ce processus. L'autocorrélogramme nous montre clairement la présence des variations saisonnières dont la période est de 12 mois. Ce processus n'est par conséquent pas stationnaire.

FIG. 4.19 Courbe des débits moyens mensuels



On pourra ainsi se servir d'un modèle $SARIMA(p,d,q)(P,D,Q)_{12}$ pour caractériser ce processus. En vue de déterminer l'ordre D, nous étudions le processus $Z_t = (1-B^{12})Y_t$. Les corrélogrammes présentés sur la figure 4.21 obtenue en utilisant la commande C.43 en annexe, nous permettent de penser que le processus Z_t est stationnaire. Ce résultat est confirmé par les tests de non stationnarité de Phillips-Perron pour lequel la probabilité critique est de 0.01. Voir commande C.44 en annexe pour la réalisation de ce test.

FIG. 4.20 Autocorrélogrammes du processus Y_t

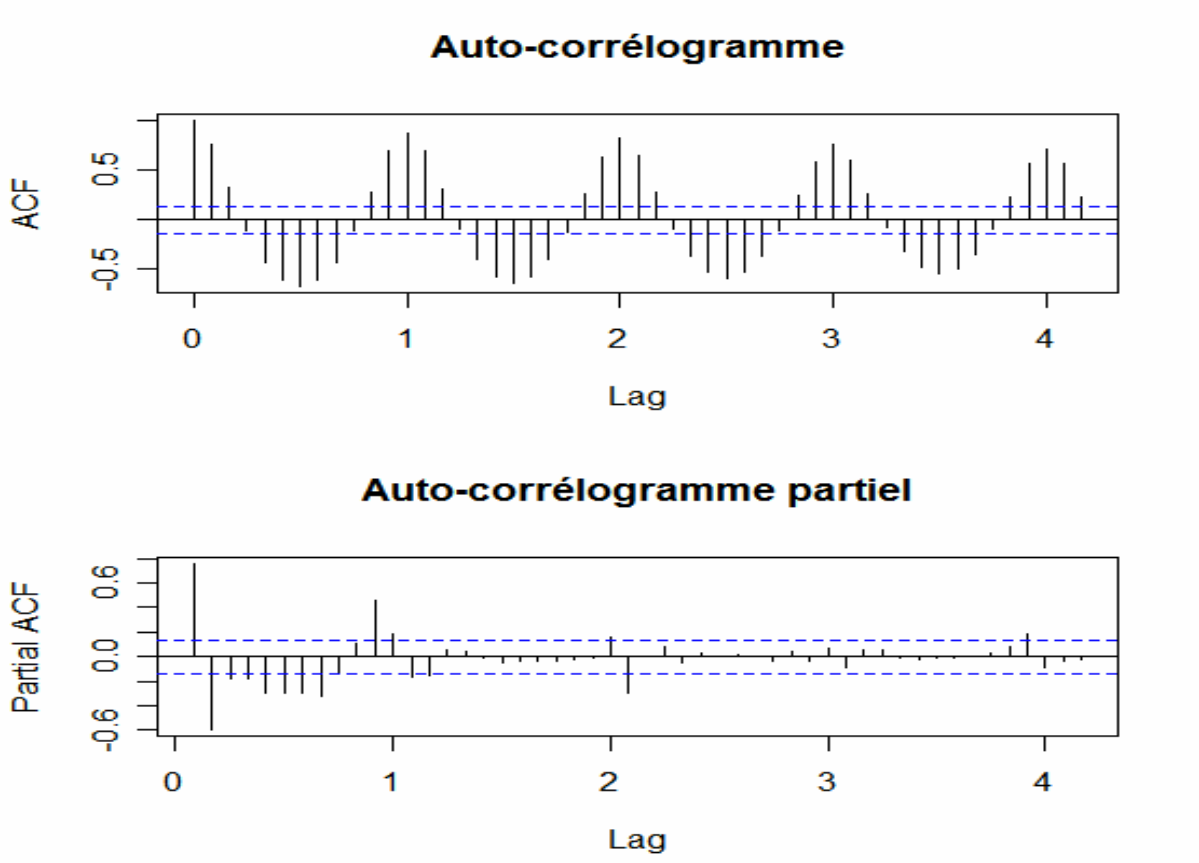
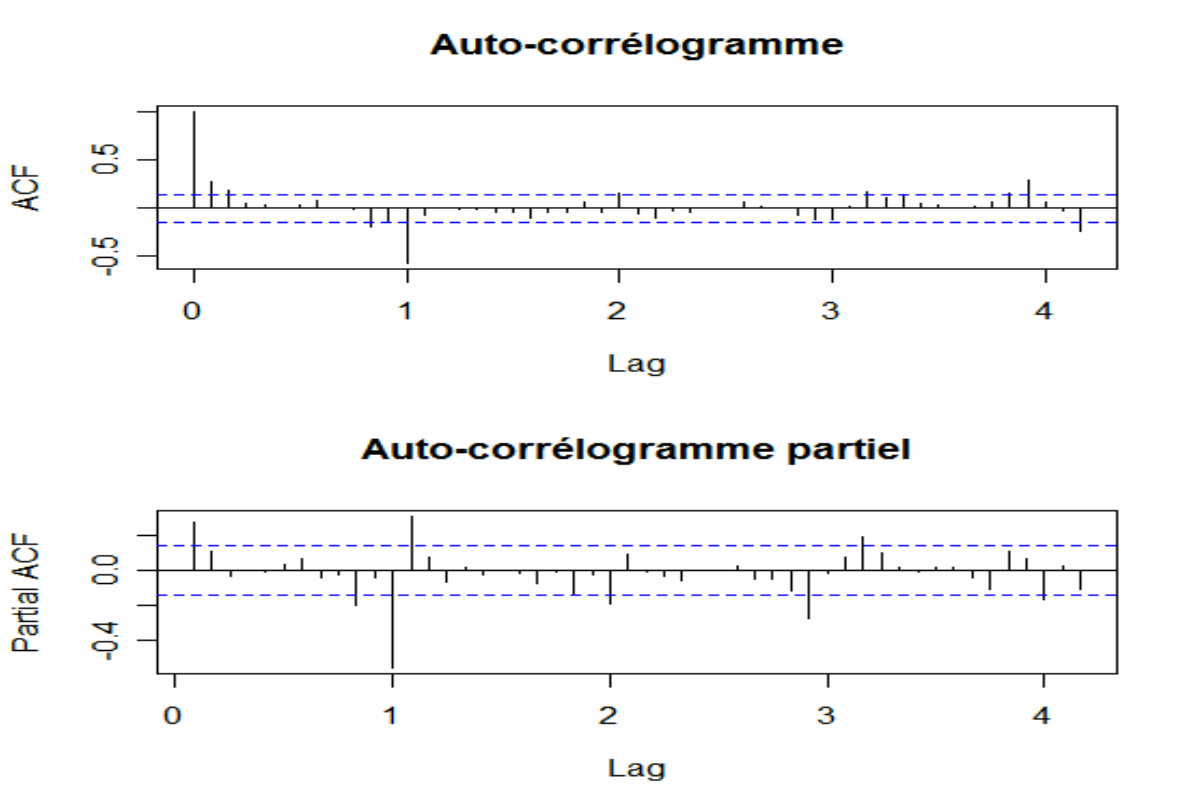


FIG. 4.21 Autocorrélogrammes du processus Z_t



Le processus Z_t étant stationnaire, on prendra alors $D = 1$. Par ailleurs, puisque nous n'avons pas eu besoin de différencier le processus Y_t dans le but d'en éliminer éventuellement une variation tendancielle linéaire et significative, on prendra alors $d = 0$.

Si nous observons la fonction d'auto corrélation du processus Z_t (figure 4.21), dans l'intervalle $[0,12]$, c'est-à-dire dans le premier intervalle de longueur égale à la période du processus Y_t, on remarque qu'elle ressemble à celle d'un processus qui suit un $MA(2)$. Par ailleurs, la fonction d'auto corrélation partielle du processus Z_t vu dans l'intervalle $[1,12]$ ressemble à celle d'un processus qui suit un $AR(1)$. Cela nous suggère alors de prendre $p = 1$ et $q = 2$.

Nous choisissons $P = 1$ et $Q = 1$. En effet, nous avons remarqué, après plusieurs estimations des paramètres de ce modèle dans lequel les ordres P et Q sont strictement supérieurs à 1, que les coefficients les des monômes du plus haut degré des polynômes de degré P et Q sont non significatifs.

L'équation générale du modèle $SARIMA(1,0,2)(1,1,1)_{12}$ est

$$\left(1-B^{12}\right)\left(1-ar1B\right)\left(1-sar1B^{12}\right)Y_t = \left(1+ma1B+ma2B^2\right)\left(1+sma1B^{12}\right)\varepsilon_t,$$

où ε_t est un bruit blanc de variance σ^2, B^p est l'opérateur retard d'ordre p et les paramètres estimés de ce modèle, ainsi que les intervalles de confiance correspondants sont stockés dans le tableau 4.10 ci-dessous . Voir commandes C.45 et C.46 en annexe pour son obtention. On note à partir de ce tableau que les coefficients ar1 et ma2 sont non significatifs. Ce qui suggère d'essayer un modèle $SARIMA(1,0,1)(1,1,1)_{12}$.

TAB. 4.10 Paramètres estimés du modèle $SARIMA(1,0,2)(1,1,1)_{12}$

	2.5 %		97.5 %
ar1	-0.0298	0.5575	1.1449
ma1	-0.8594	-0.2602	0.3389
ma2	-0.2302	0.0343	0.2989
sar1	-0.3875	-0.2466	-0.1056
sma1	-1.1901	-1.000	-0.8099
σ^2		142758	

Les paramètres du modèle $SARIMA(1,0,1)(1,1,1)_{12}$ ainsi que les intervalles de confiance correspondants sont stockés dans le tableau 4.11. Voir les commandes C.47 et C.48 dans l'annexe pour son obtention. On note surtout dans ce tableau la non significativité du coefficient ma1. Ce qui suggère d'essayer un modèle $SARIMA(1,0,0)(1,1,1)_{12}$.

TAB. 4.11 Paramètres estimés du modèle $SARIMA(1,0,1)(1,1,1)_{12}$

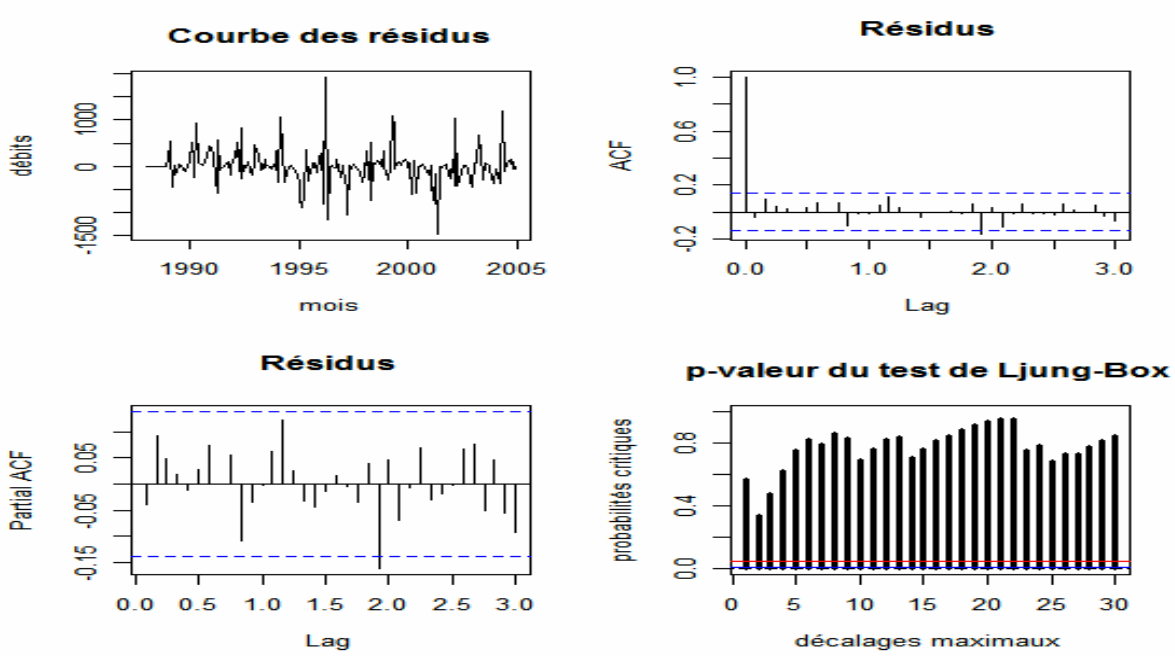
	2.5 %		97.5 %
ar1	0.3218	0.6201	0.9185
ma1	-0.6778	-0.3201	0.0375
sar1	-0.3881	-0.2475	-0.1069
sma1	-1.1897	-0.9999	-0.8102
σ^2		142797	

Le tableau 4.12 contient les paramètres estimés du modèle $SARIMA(1,0,0)(1,1,1)_{12}$ ainsi que les intervalles de confiance correspondants. Voir les commandes C.49 et C.50 dans l'annexe pour son obtention. On note sur ce tableau la significativité de tous les coefficients du modèle.

TAB. 4.12 Paramètres estimés du modèle $SARIMA(1,0,0)(1,1,1)_{12}$

	2.5 %		97.5 %
ar1	0.2043	0.3381	0.4719
sar1	-0.3849	-0.2441	-0.1032
sma1	-1.1926	-0.9999	-0.8074
σ^2		144849	

FIG. 4.22 Diagnostic des résidus ε_t



Vérifions à présent l'hypothèse de bruit blanc des résidus du modèle $SARIMA(1,0,0)(1,1,1)_{12}$. On observe sur le graphes de la figure 4.22 ci-dessus obtenue en utilisant la commande C.51 en annexe, que ces résidus forment effectivement un bruit blanc.

Conclusion générale de cette modélisation

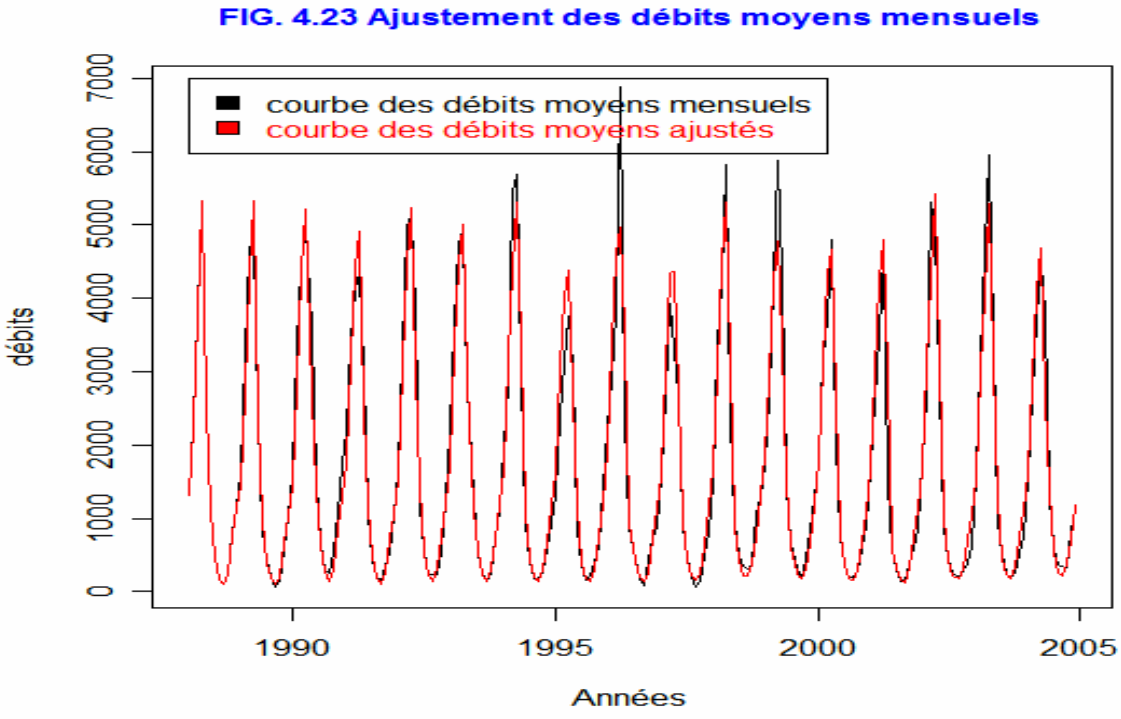
Nous terminons cette partie en disant que l'équation finale estimée du modèle vérifié par la chronique de nos débits moyens mensuels Y_t est

$$(1-B^{12})(1-0,3381B)(1+0,2441B^{12})Y_t = (1-0.9999B^{12})\varepsilon_t,$$

où ε_t est un bruit blanc de variance estimée $\hat{\sigma}^2 = 144849$.

Sur la figure 4.23 obtenue en utilisant la commande C.52 dans l'annexe, on peut voir la courbe des débits moyens mensuels sur laquelle est superposée celle ajustée par ce modèle. Le tableau 4.13 contient les erreurs quadratiques relatives à cet ajustement. Voir commande C.53 en annexe pour l'obtention de ce tableau.

FIG. 4.23 Ajustement des débits moyens mensuels



TAB. 4.13 Erreurs quadratiques de l'ajustement du modèle aux débits moyens mensuels Y_t

	Nombre de jours	er (%)	em (m^3/s)
Ajustement	204	15.4	369.2

Notons que la prévision des débits moyens mensuels ne présente pas un grand intérêt pour AES-SONEL, étant donné qu'il lui est plus important de connaître les prévisions journalières. Nous pouvons continuer cette dernière modélisation en cherchant à prédire les débits naturels journaliers à l'intérieur des différents mois sous la contrainte que les débits moyens de ces mois sont connus. Mais le temps qui nous a été imparti (durée du stage) ne nous a pas permis de pousser cette analyse un peu plus loin.

Nous terminons ce chapitre en concluant que la modélisation des débits naturels journaliers du B.V.I. à Sondmbengué par un modèle faisant intervenir un processus de type ARIMA après une standardisation de cette chronique de débits ne semble pas être différente de la modélisation de ces mêmes débits par un modèle multiplicatif après une désaisonnalisation de cette chronique par la méthode des moyennes mobiles, dans la mesure où les erreurs quadratiques de l'ajustement global ainsi que les erreurs quadratiques de la prévision a posteriori de ces deux modèles aux données réelles sont quasiment équivalentes. Ainsi, nous avons écrit un programme R qui permet de faire des prévisions des débits naturels journaliers de la Sanaga à la station hydrologique de Songmbengué. Ce programme est basé sur le principe de la première modélisation. Voir commande C.54 dans l'annexe pour le code de ce programme.

Par ailleurs, nous avons ajouté à cette étude un programme R permettant de faire des simulations des débits naturels journaliers de la Sanaga à la station de contrôle de Songbemgué. Ce programme est basé sur un principe markovien du premier ordre décrit à la section 3.5 du chapitre 3. Voir commande C.55 en annexe pour le code de ce programme. Nous présentons sur la figure 4.24 et sur la figure 4.25 les courbes des débits naturels journaliers obtenus par deux simulations distinctes à l'aide de ce programme de simulation. Notons que nous avons simulé les débits naturels de 17 années soit de 6205 jours et que la première valeur simulée est celle du 1er Juillet. Voir commande C.56 dans l'annexe pour la réalisation de ces deux simulations.

FIG. 4.24 Courbe des débits journaliers simulés

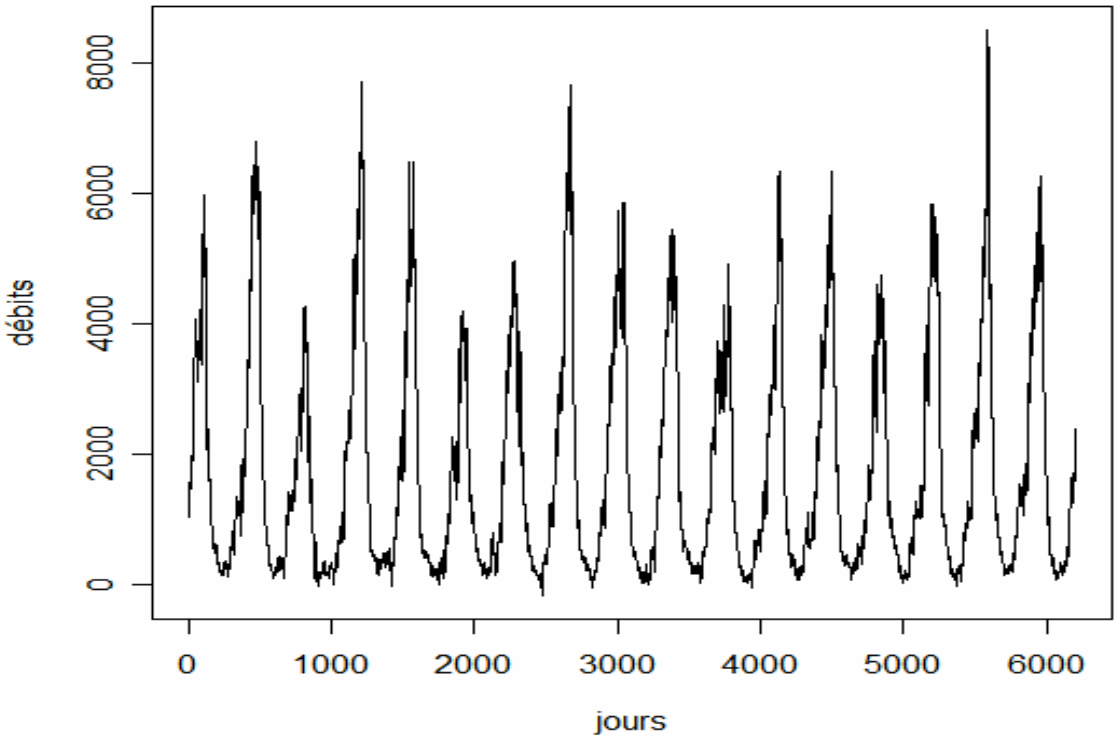
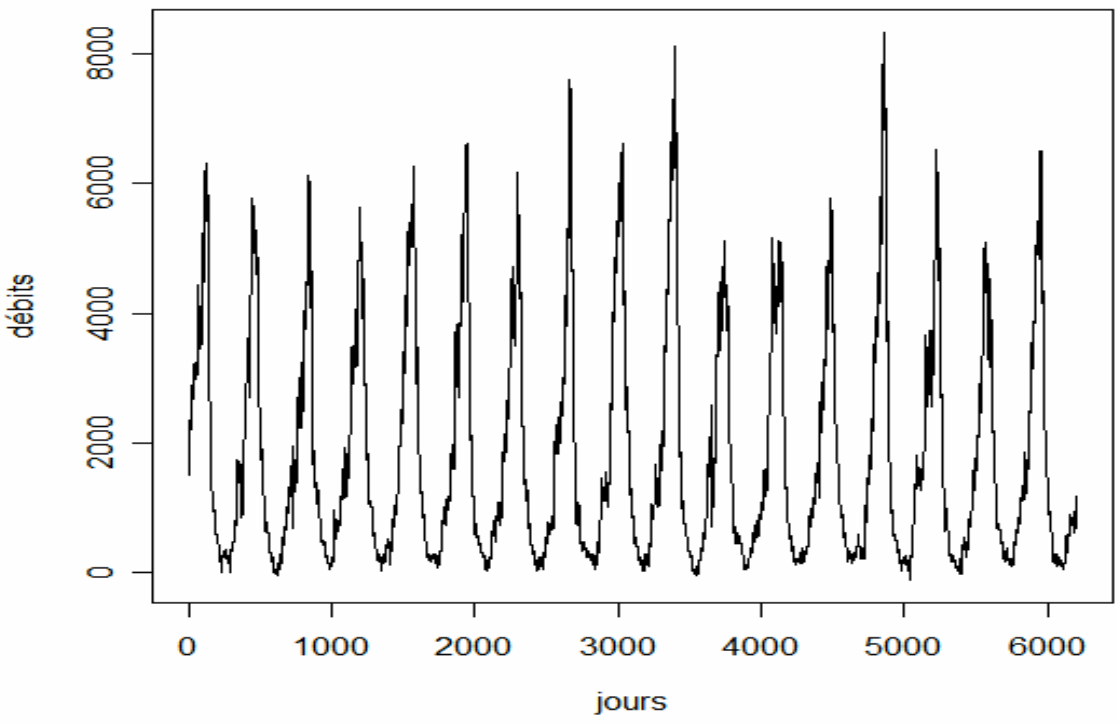


FIG. 4.25 Courbe des débits journaliers simulés



Conclusion générale

Dans le but de maîtriser le coût de la dépense en volume de combustibles nécessaires pour la production d'énergie à partir des centrales thermiques et de planifier l'exploitation courante du système hydroélectrique en vue de répondre aux besoins de la production, la détermination d'un modèle de prévision des débits naturels journaliers du B.V.I. de la Sanaga à la station de contrôle de Songmbengué est recommandée.

Pour y parvenir, nous avons établi deux modèles de prévision. Le premier modèle de prévision est basé sur un processus de type $ARIMA(11,1,1)$ appliqué à la série des débits naturels journaliers standardisés. Le second modèle de prévision est basé sur un modèle de type multiplicatif appliqué à ces débits et, ceci après une désaisonnalisation de cette série par la méthode des moyennes mobiles. Les modèles de prévision obtenus au terme de notre étude ont été jugés satisfaisants et équivalents en terme d'erreurs quadratiques d'ajustement et de prévision a posteriori. Par ailleurs, un modèle de simulation a été mis au point.

Les erreurs quadratiques de prévision obtenues dans cette étude ne sont pas assez faibles. Nous proposons la démarche suivante pour corriger dans l'avenir les prévisions des débits naturels journaliers d'une année k (k>2005).

1- premièrement, soumettre la série des erreurs de prévision à posteriori obtenue au cours de la période allant du 1er juillet 2001 au 30 juin de l'année k-1 à une analyse chronologique;

2- deuxièmement, prédire les erreurs de prévisions à posteriori de l'année k à partir du modèle obtenu pour ces erreurs de prévision;

3- troisièmement, prédire les débits naturels journaliers de l'année k à partir de l'un des deux modèles obtenu dans cette étude;

4- quatrièmement, corriger ces prévisions des débits naturels journaliers de l'année k en leurs ajoutant les erreurs de prédiction à posteriori prédites pour l'année k.

Annexe

TAB. 2.2 Série des débits moyens mensuels du B.V.I. de la Sanaga à la station de contrôle de Songmbengué

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1988							1320,81	2272,00	3789,03	5332,13	2510,17	1107,19
1989	457,71	179,50	102,13	241,17	794,13	1143,33	1602,32	3246,94	4638,40	4876,35	2395,17	839,26
1990	431,10	222,04	59,03	135,97	579,52	1021,63	1590,16	3351,10	4439,33	4978,74	3582,20	1506,03
1991	675,16	290,18	249,10	541,33	1068,23	1723,47	2187,58	3416,16	4131,03	4347,35	3021,57	1074,39
1992	450,55	204,64	136,29	340,00	534,45	1053,93	1645,19	3243,16	4987,00	5117,39	3818,60	1177,81
1993	565,03	217,04	235,77	232,90	562,03	1264,73	2194,26	3647,35	4735,27	4928,10	2821,03	1275,16
1994	516,00	245,36	137,35	250,40	748,52	1110,13	1973,74	2845,26	5451,50	5686,71	2985,80	935,87
1995	422,61	183,86	176,29	271,23	588,23	944,13	1308,39	2381,19	3138,53	3753,35	3068,47	860,74
1996	359,90	150,86	213,55	401,03	611,65	1406,50	2223,13	2905,71	3812,20	6886,71	2389,00	900,87
1997	445,97	178,29	87,03	420,13	743,13	882,87	1600,10	2702,19	3935,83	3295,97	2697,87	921,81
1998	469,74	203,61	73,16	136,17	491,84	806,97	1493,55	3351,45	4206,50	5822,65	2379,03	959,74
1999	476,84	349,25	313,52	371,00	1060,52	1284,97	1739,90	2643,61	4222,70	5879,16	4309,80	1447,65
2000	658,26	336,07	182,03	443,50	819,32	1282,20	1776,71	3199,10	3714,63	4790,42	2047,17	763,77
2001	364,81	176,36	188,81	269,07	602,19	901,43	1696,90	2610,23	3495,27	4635,61	1492,03	574,23
2002	305,32	137,86	155,84	363,17	587,52	867,73	1791,84	2667,29	5315,37	4990,45	2833,83	879,39
2003	442,90	249,18	201,77	294,17	390,65	609,07	1616,87	3094,00	4709,50	5953,19	3447,30	961,35
2004	510,32	261,79	183,52	254,83	517,48	748,70	1552,97	2681,90	3717,47	4496,84	3811,90	1289,13
2005	589,35	365,32	357,71	331,77	813,81	1126,73						

Minimum	305,32	137,86	59,03	135,97	390,65	609,07	1308,39	2272,00	3138,53	3295,97	1492,03	574,23
Maximum	675,16	365,32	357,71	541,33	1068,23	1723,47	2223,13	3647,35	5451,50	6886,71	4309,80	1506,03
Moyenne	478,92	232,42	179,58	311,64	677,25	1069,32	1724,38	2956,39	4261,15	5045,36	2918,29	1027,91
Ecart-type	99,73	68,68	80,65	106,31	189,22	272,41	278,31	395,33	644,59	859,26	714,72	246,37

La série des figures ci-dessous représente l'évolution journalière des débits naturels du B.V.I. suivant un découpage mensuel.

FIG. 2.5 Débits du mois de Décembre par année

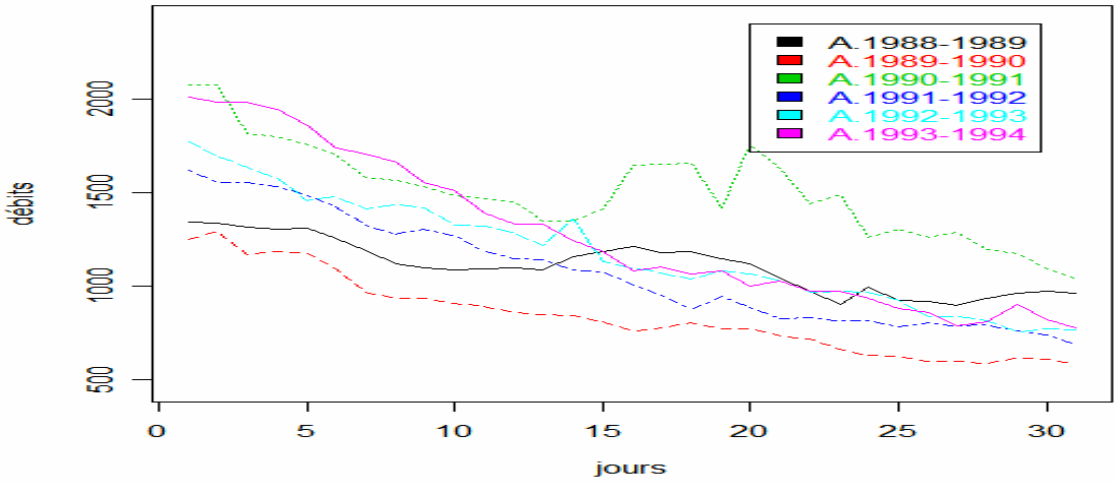


FIG. 2.6 Débits du mois de Décembre par année

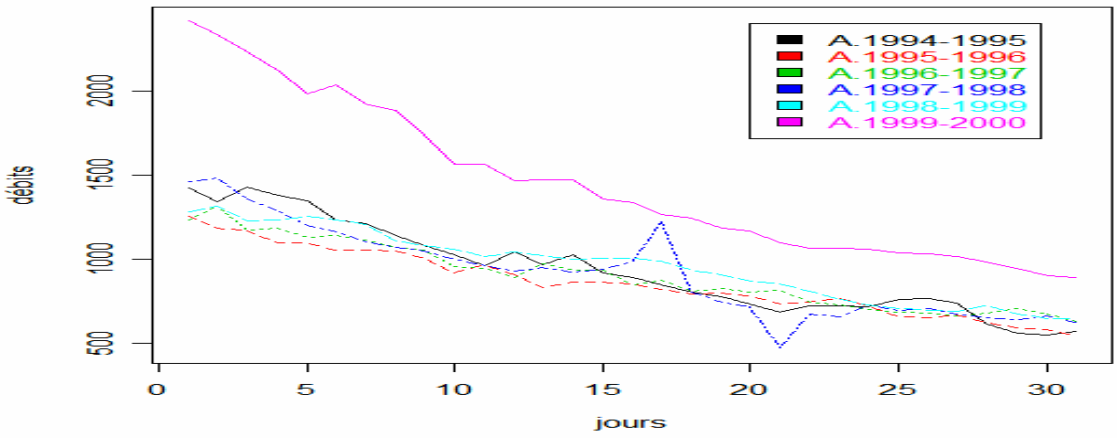


FIG. 2.7 Débits du mois de Décembre par année

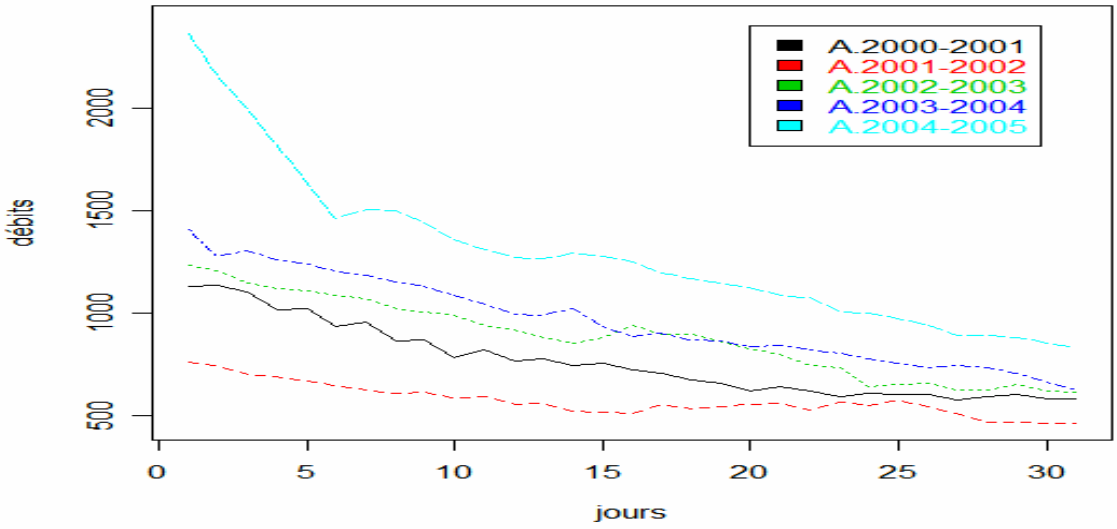


FIG. 2.8 Débits du mois de Janvier par année

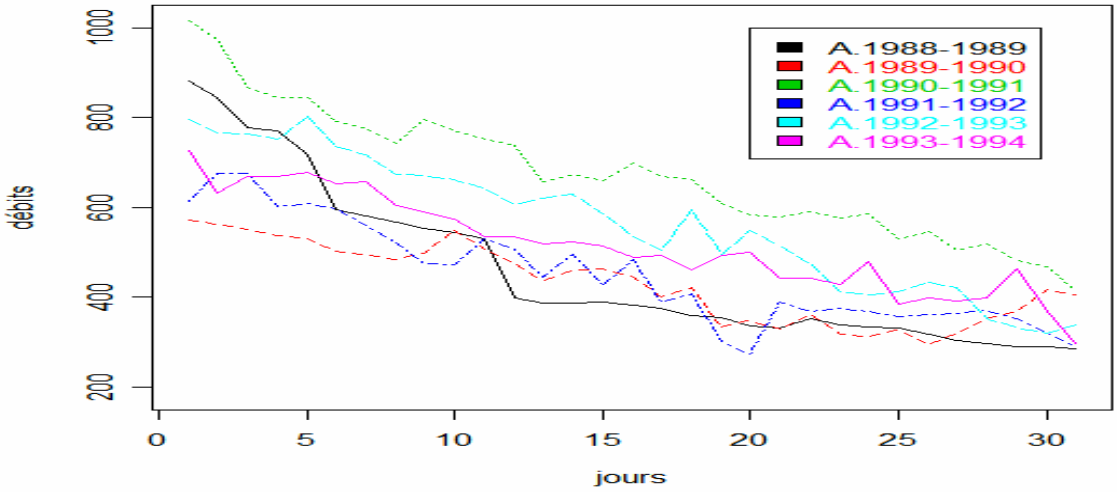


FIG. 2.9 Débits du mois de Janvier par année

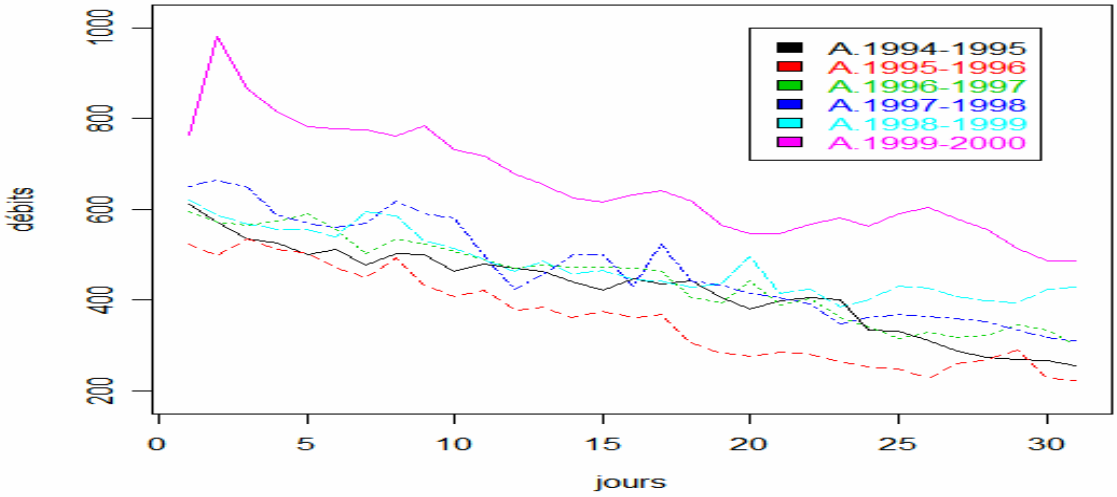


FIG. 2.10 Débits du mois de Janvier par année

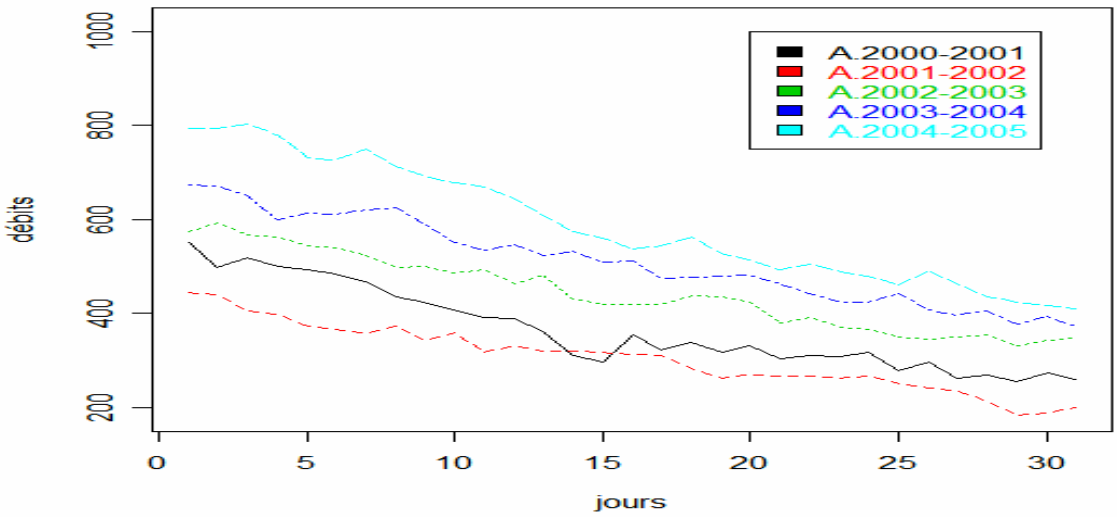


FIG. 2.11 Débits du mois de Février par année

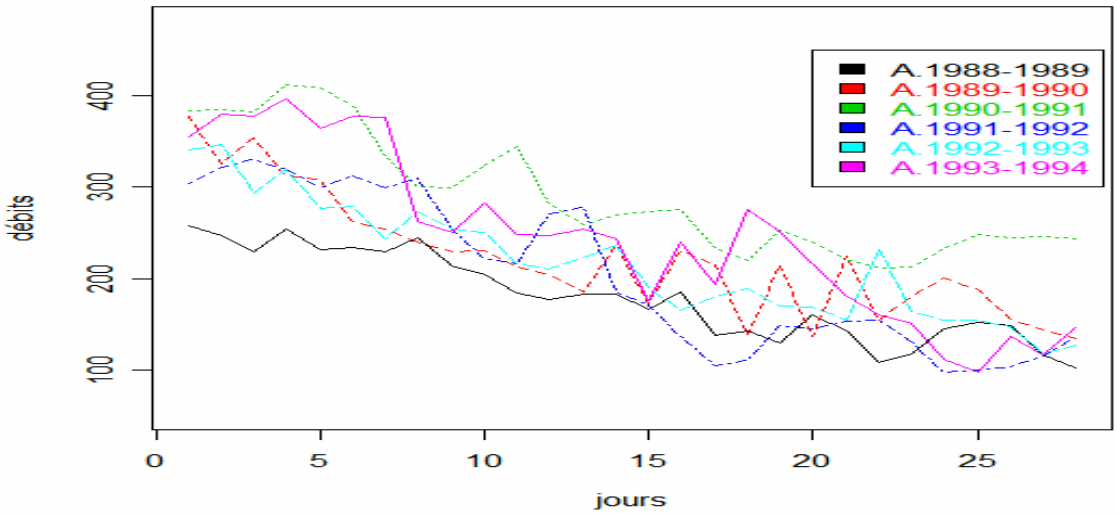


FIG. 2.12 Débits du mois de Février par année

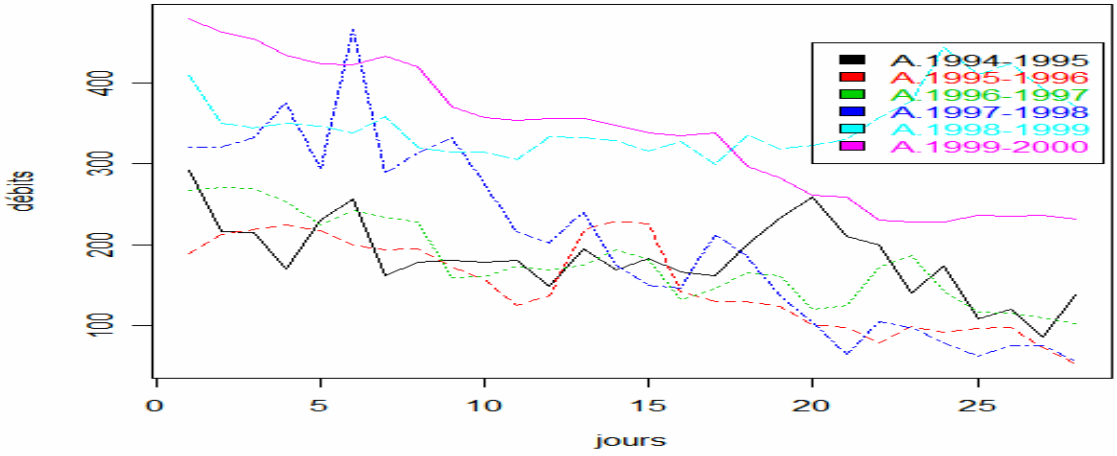


FIG. 2.13 Débits du mois de Février par année

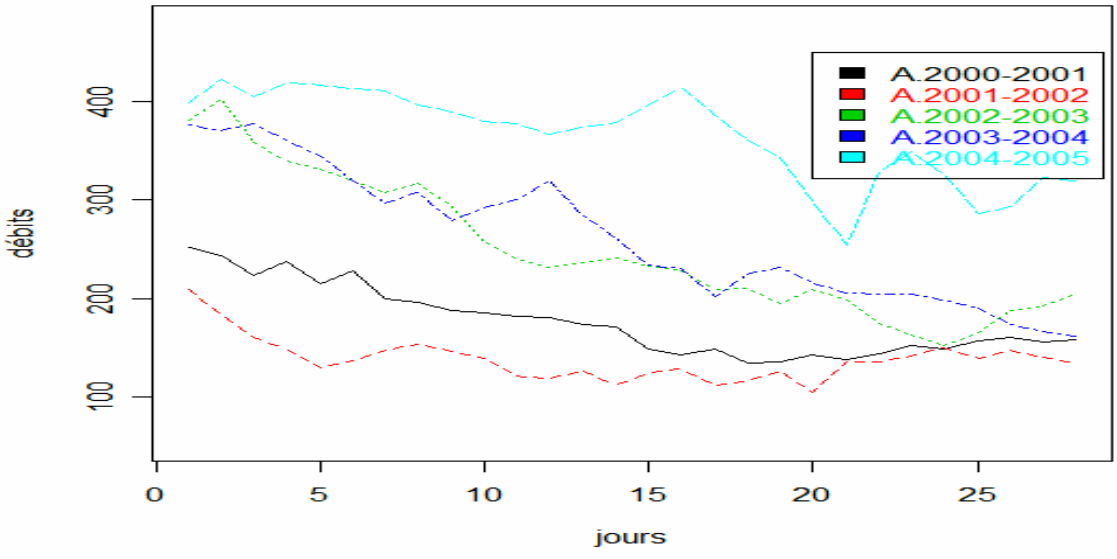


FIG. 2.14 Débits du mois de Mars par année

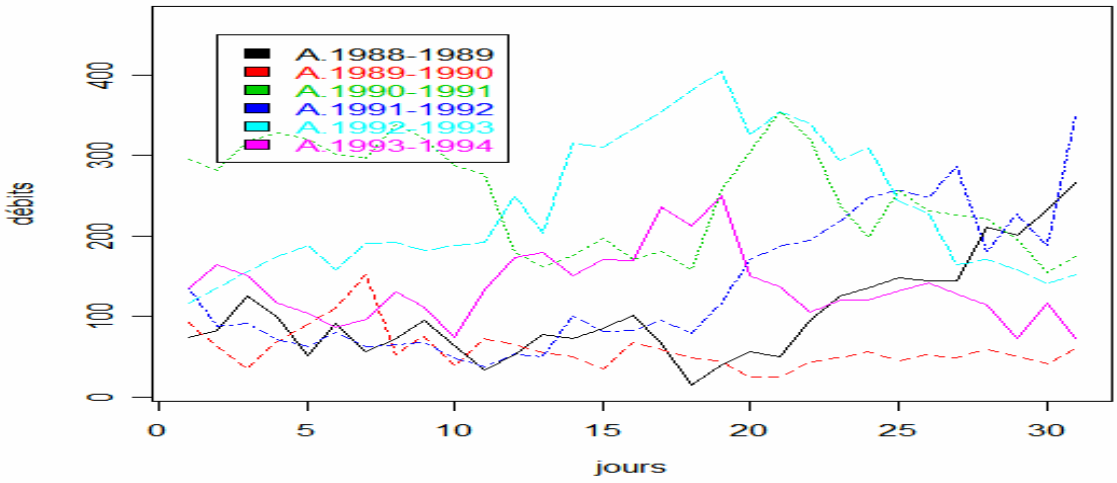


FIG. 2.15 Débits du mois de Mars par année

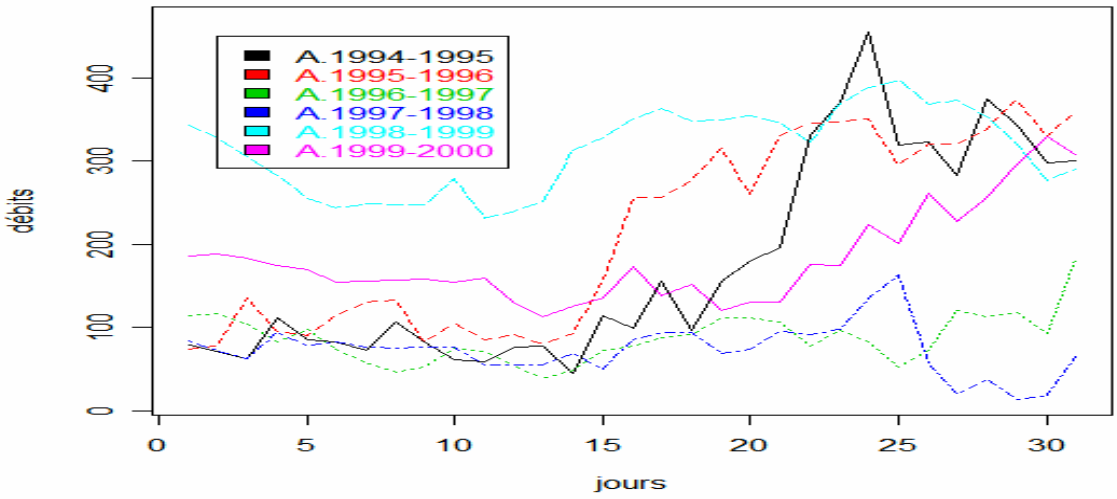


FIG. 2.16 Débits du mois de Mars par année

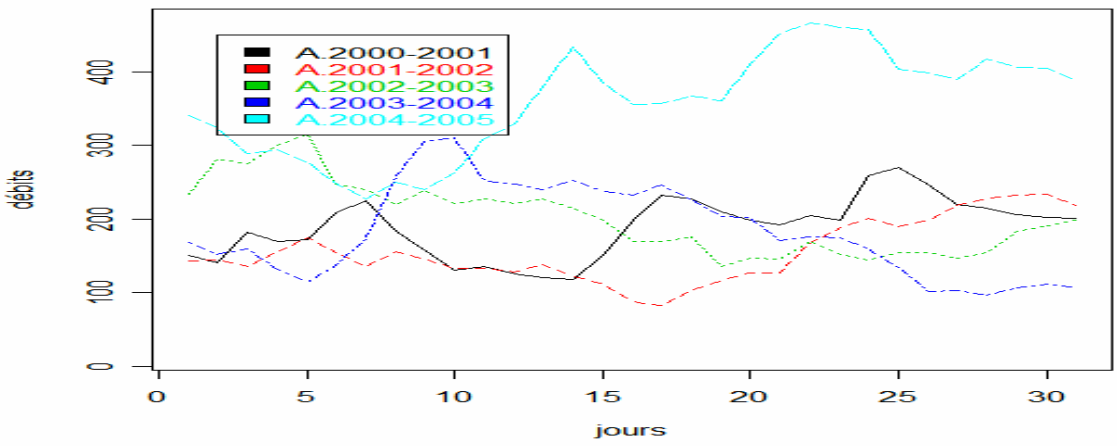


FIG. 2.17 Débits du mois d'Avril par année

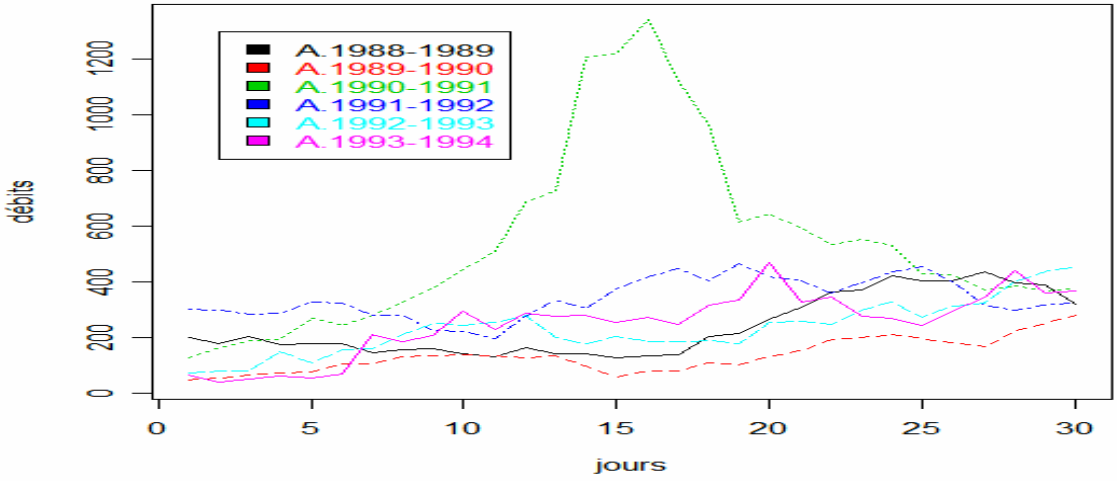


FIG. 2.18 Débits du mois d'Avril par année

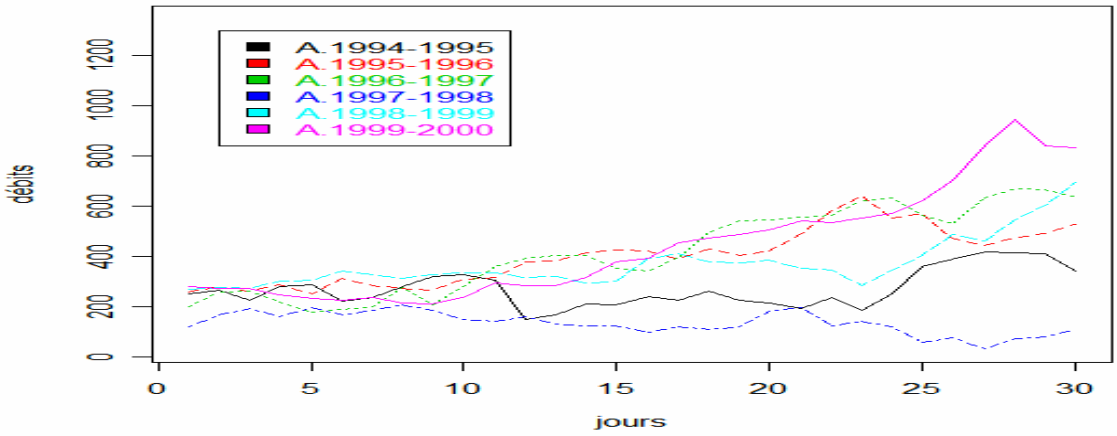


FIG. 2.19 Débits du mois d'Avril par année

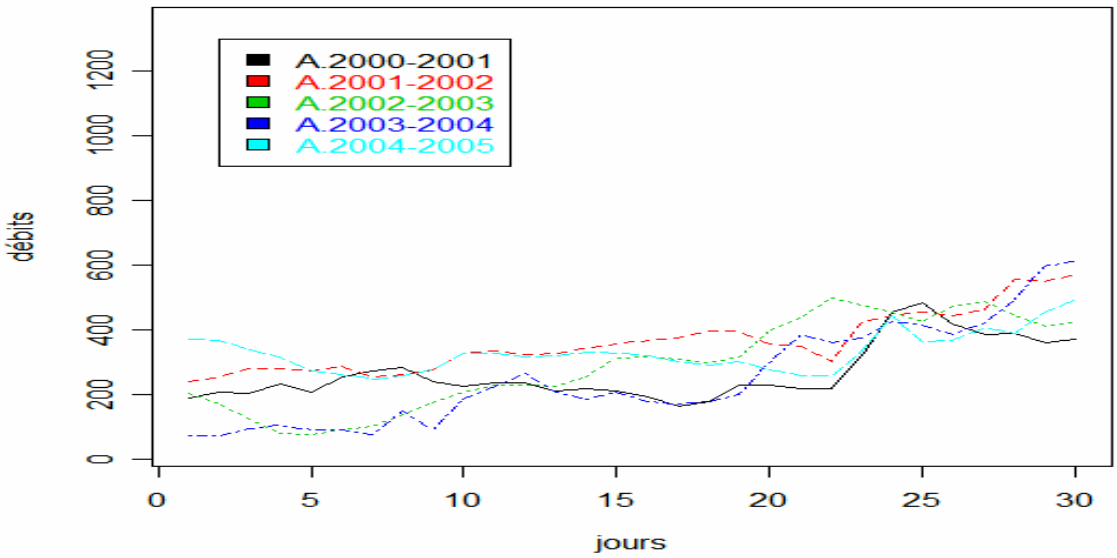


FIG. 2.20 Débits du mois de Mai par année

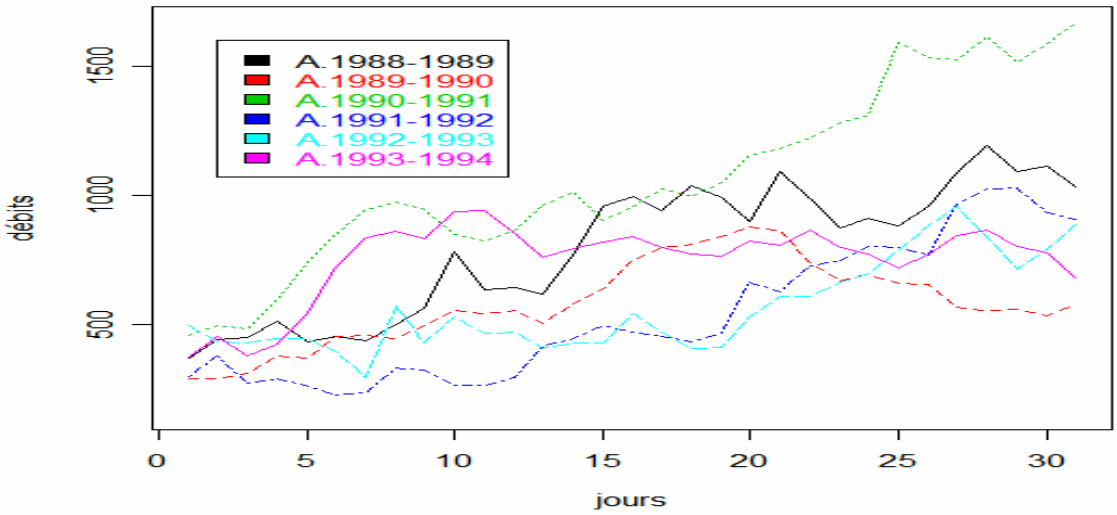


FIG. 2.21 Débits du mois de Mai par année

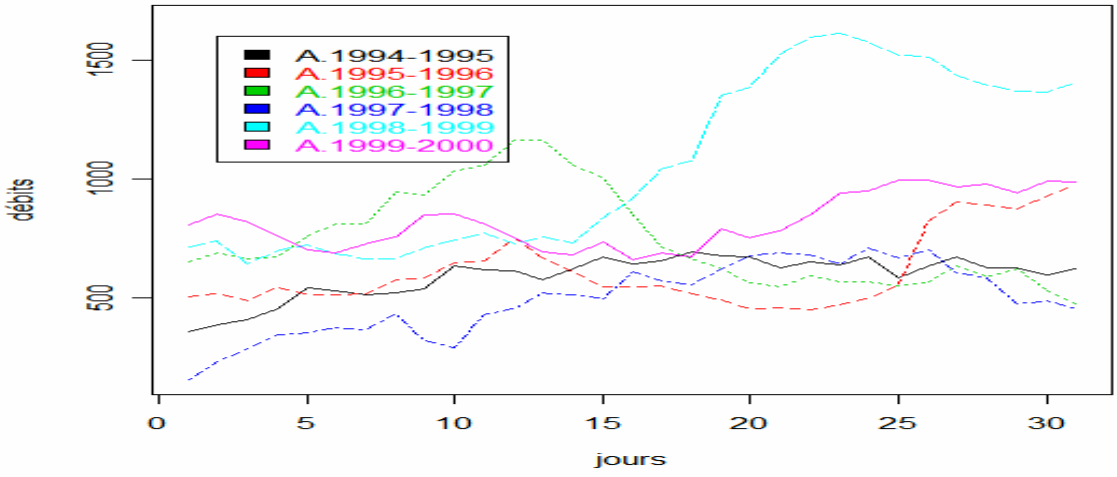


FIG. 2.22 Débits du mois de Mai par année

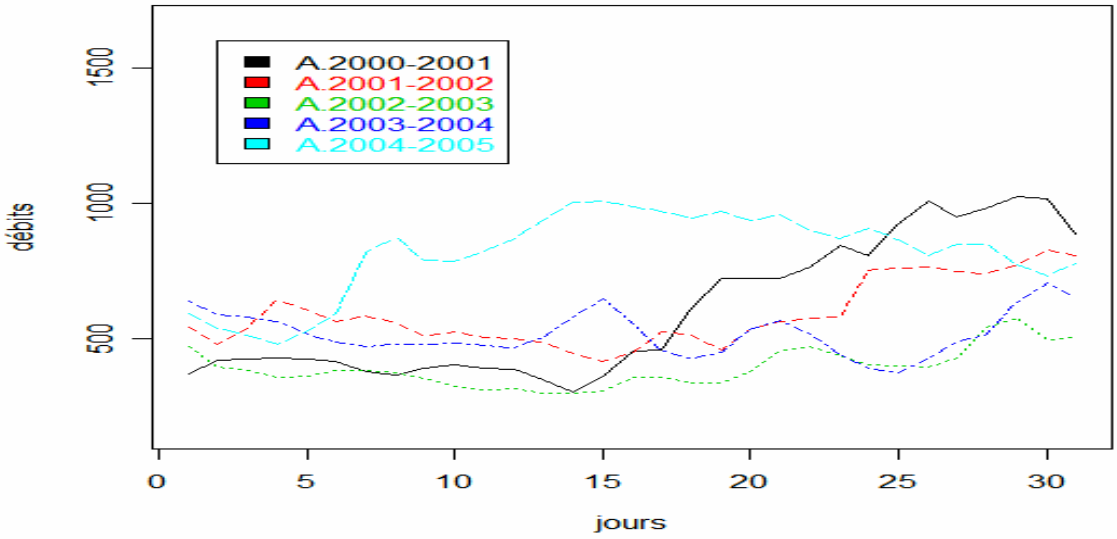


FIG. 2.23 Débits du mois de Juin par année

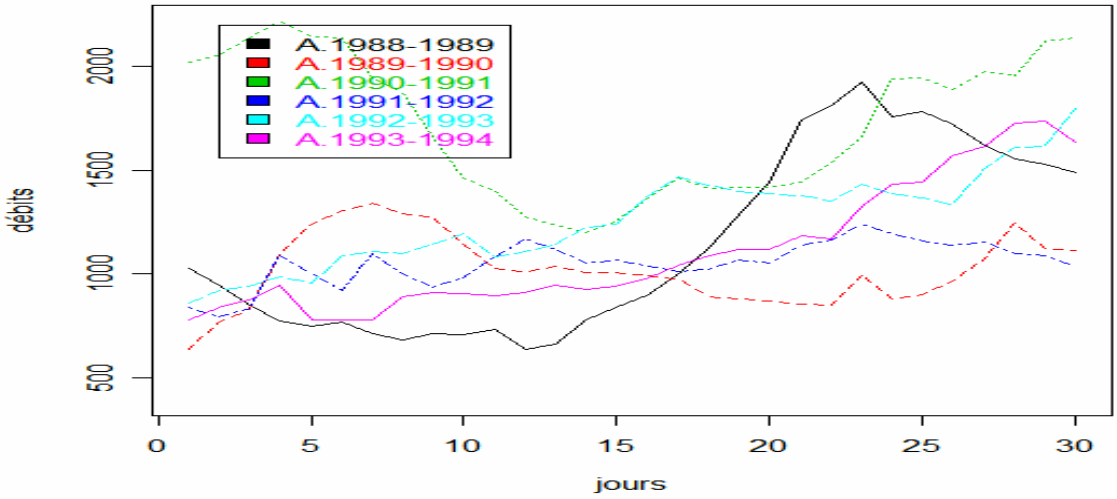


FIG. 2.24 Débits du mois de Juin par année

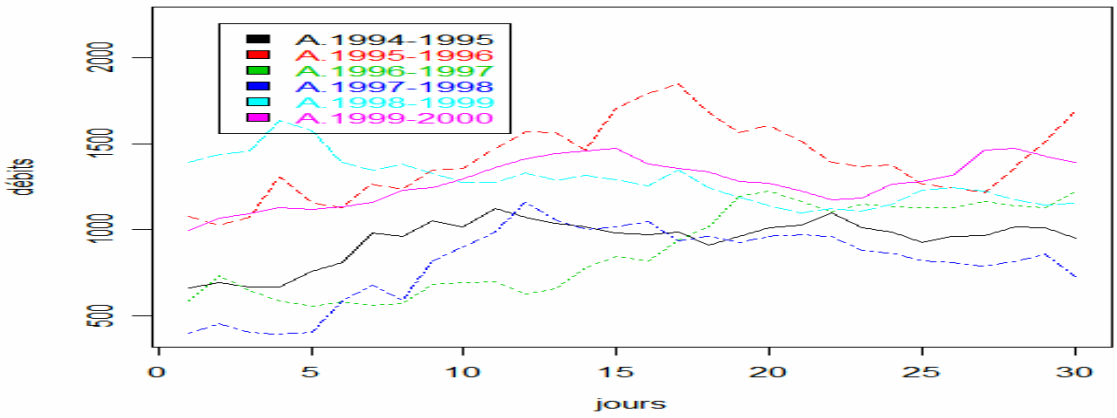


FIG. 2.25 Débits du mois de Juin par année

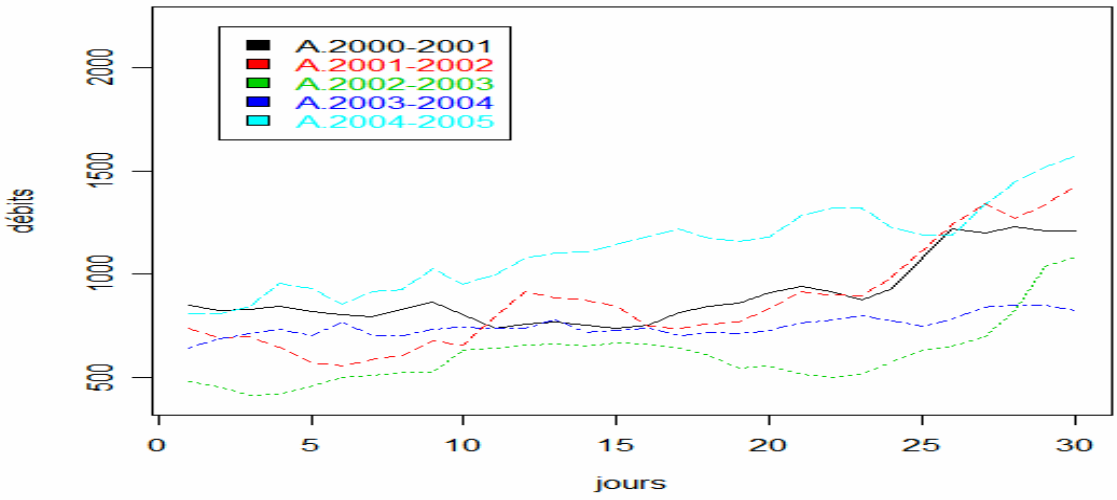


FIG. 2.26 Débits du mois de Juillet par année

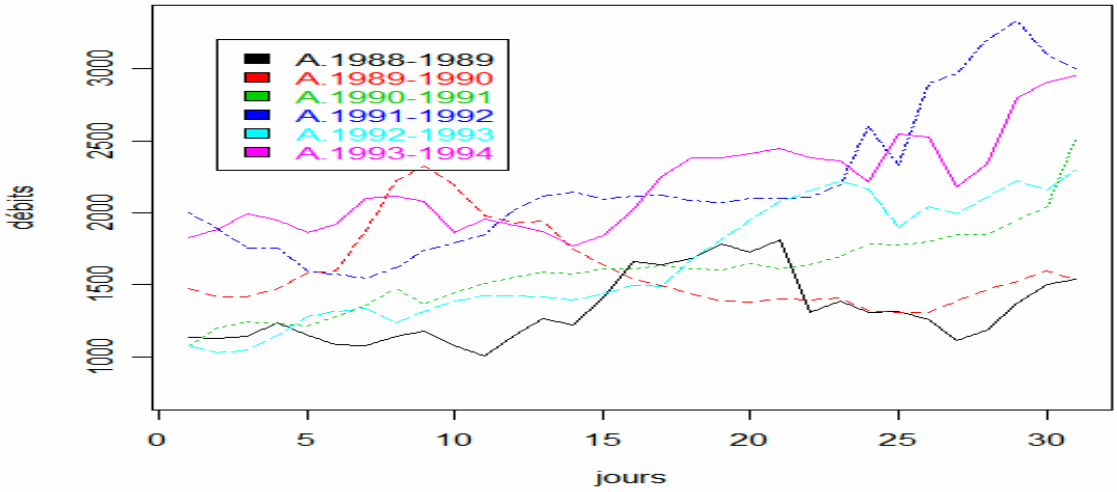


FIG. 2.27 Débits du mois de Juillet par année

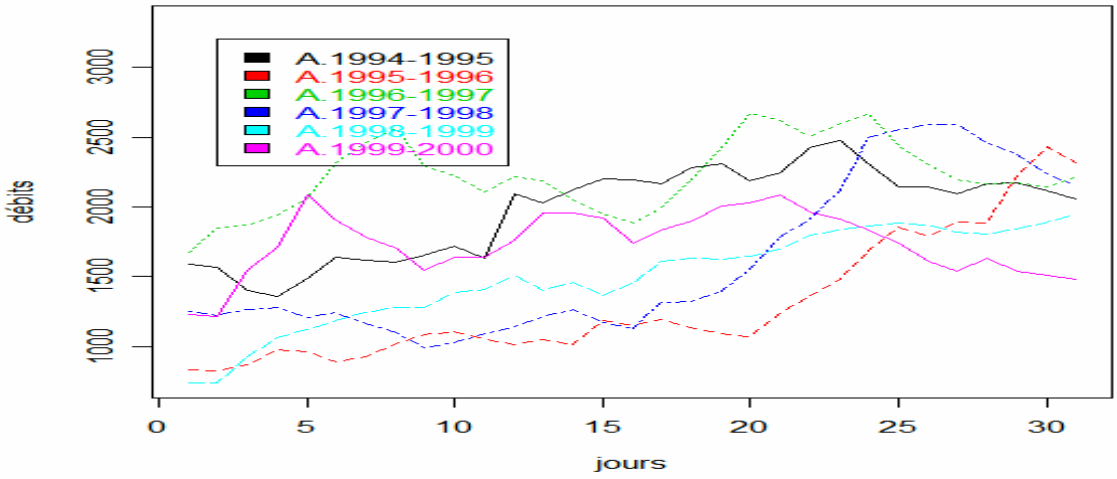


FIG. 2.28 Débits du mois de Juillet par année

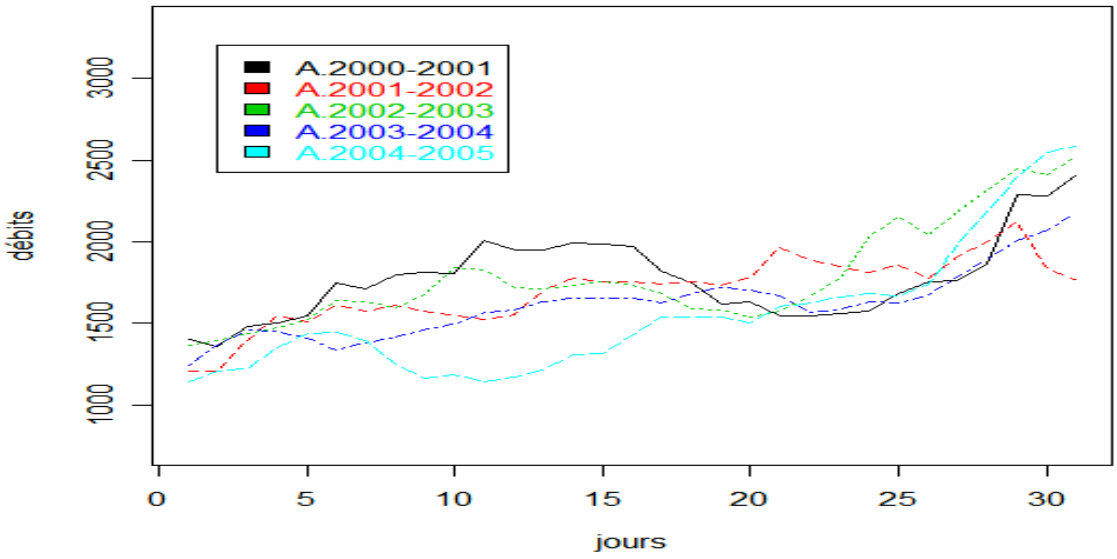


FIG. 2.29 Débits du mois d'Août par année

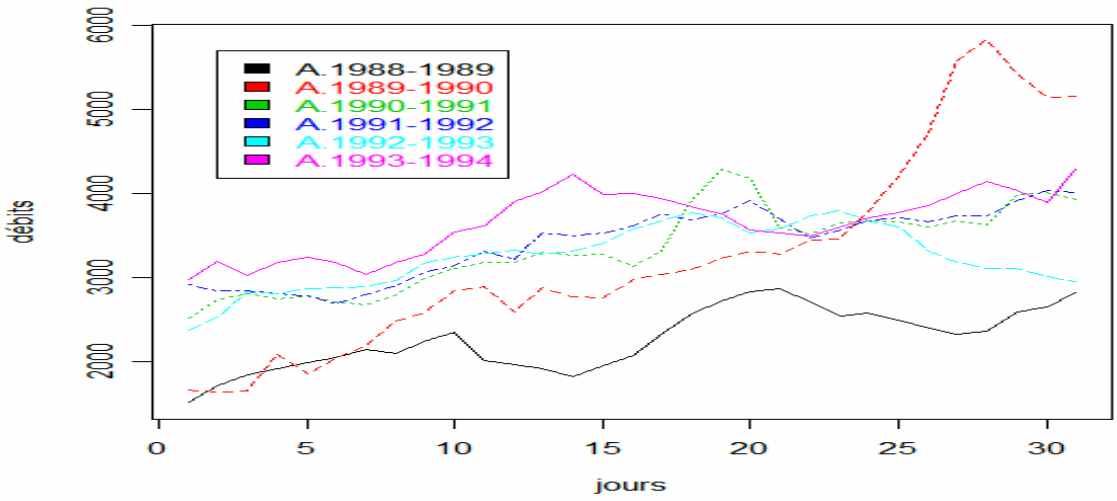


FIG. 2.30 Débits du mois d'Août par année

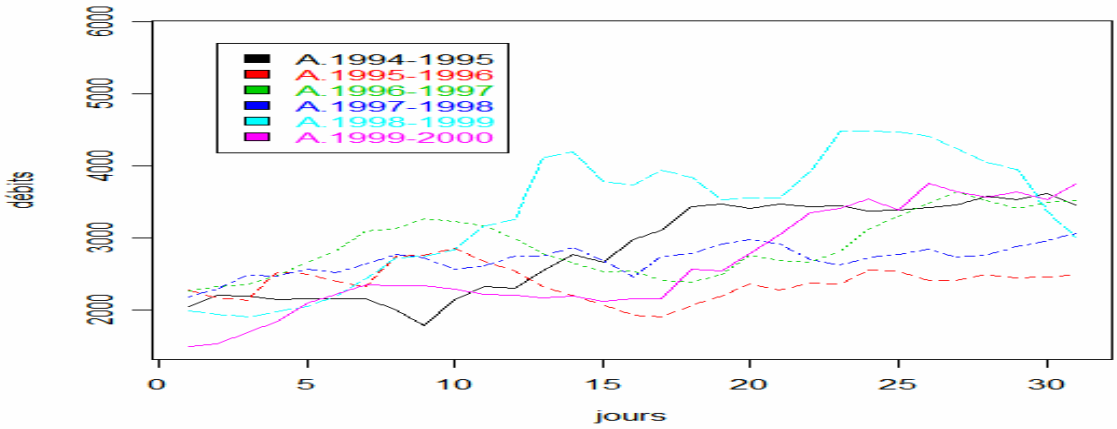


FIG. 2.31 Débits du mois d'Août par année

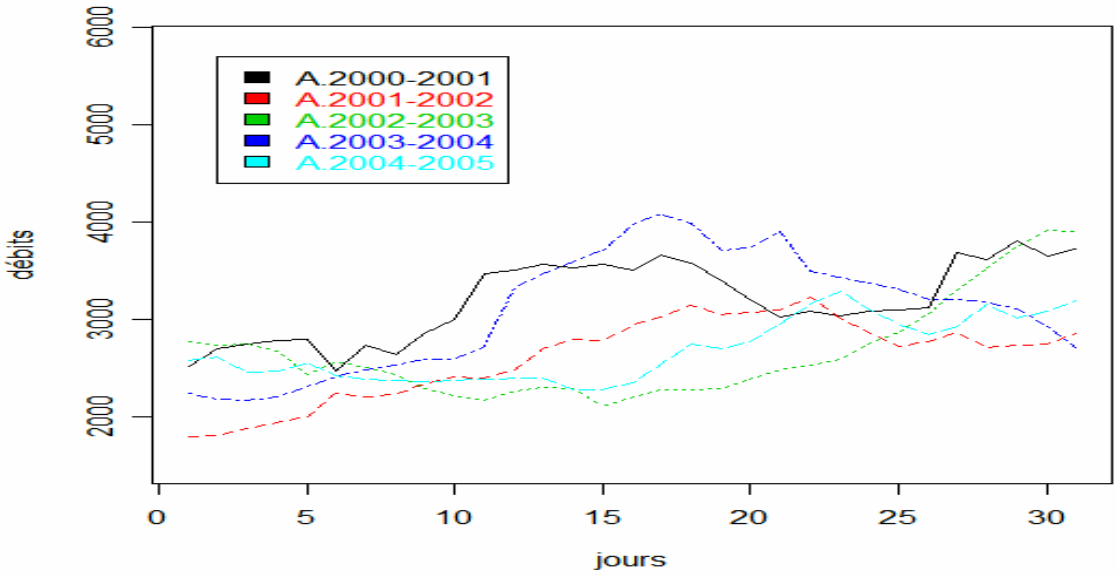


FIG. 2.32 Débits du mois de Septembre par année

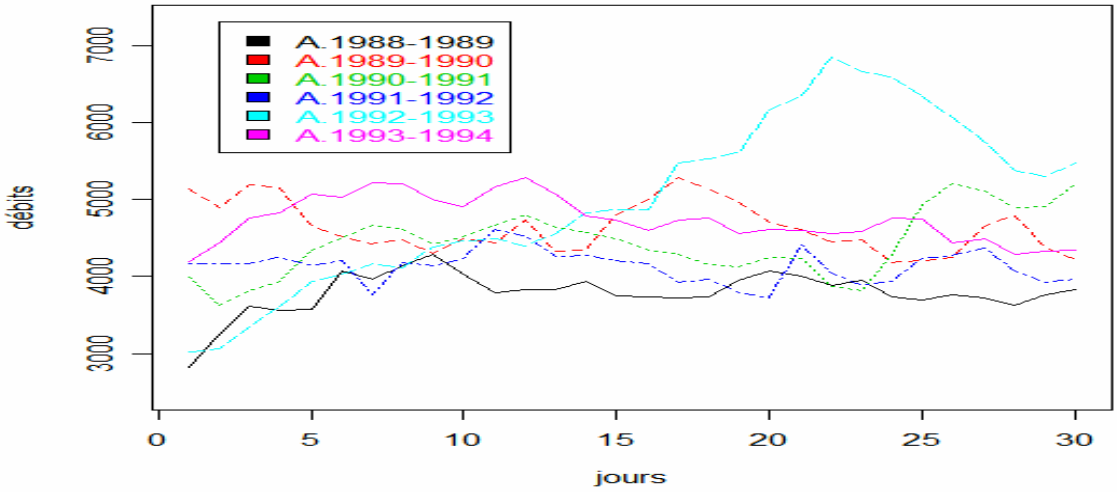


FIG. 2.33 Débits du mois de Septembre par année

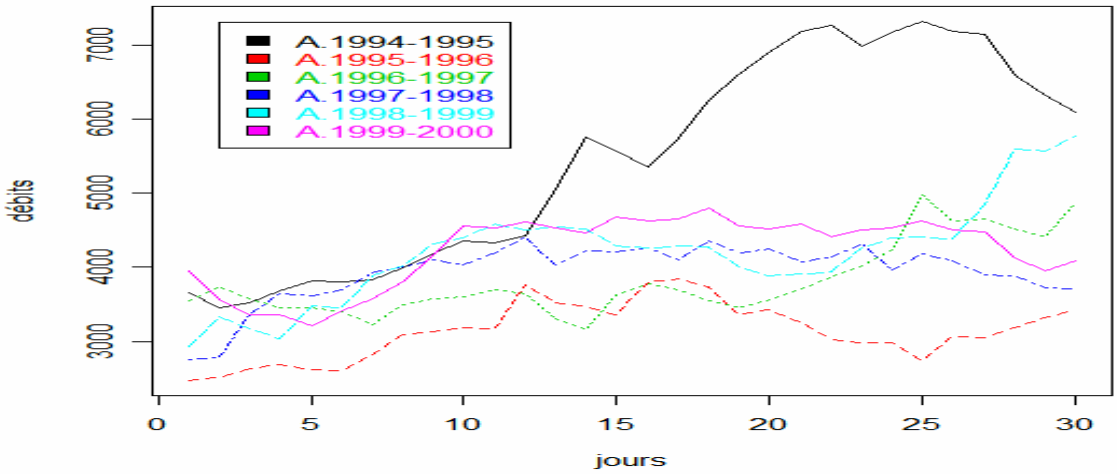


FIG. 2.34 Débits du mois de Septembre par année

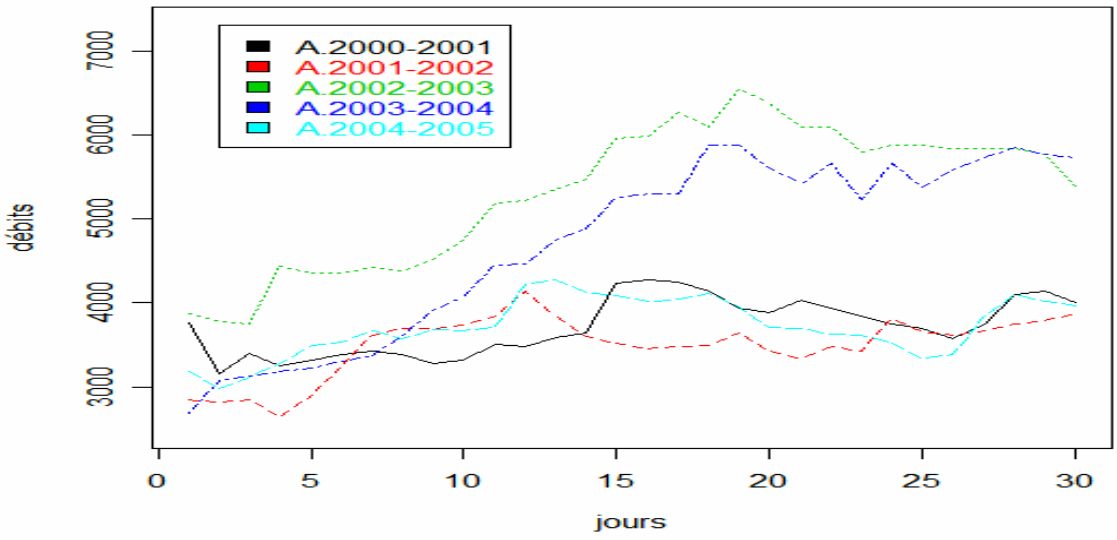


FIG. 2.35 Débits du mois d'Octobre par année

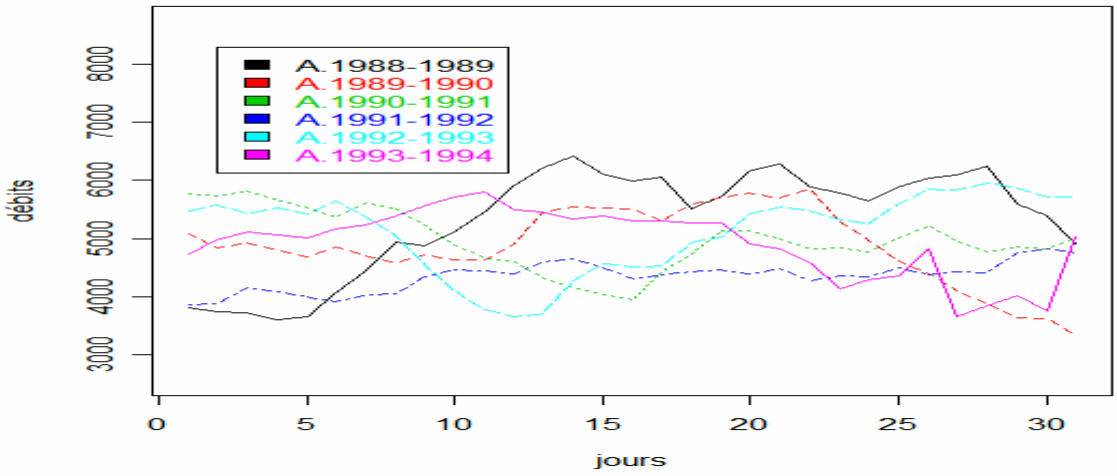


FIG. 2.36 Débits du mois d'Octobre par année

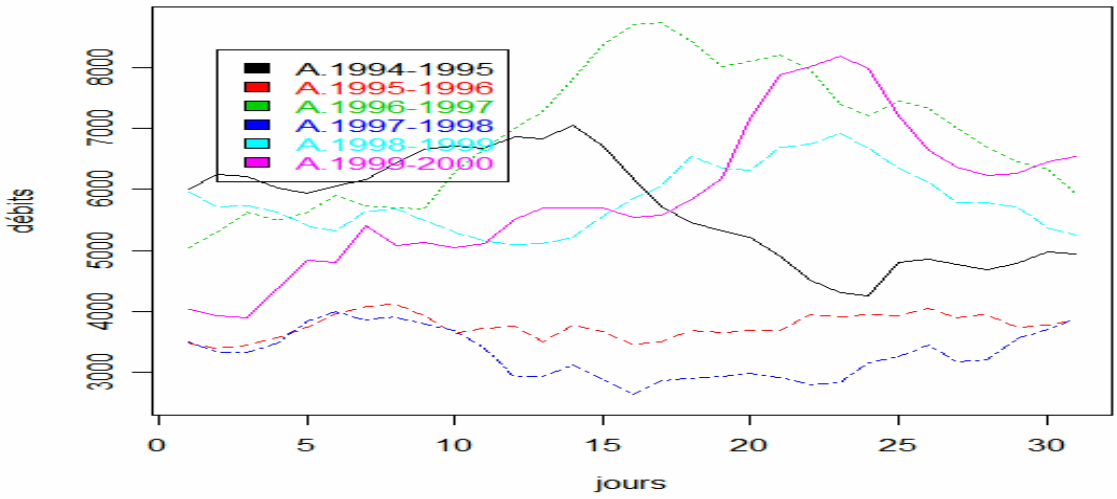


FIG. 2.37 Débits du mois d'Octobre par année

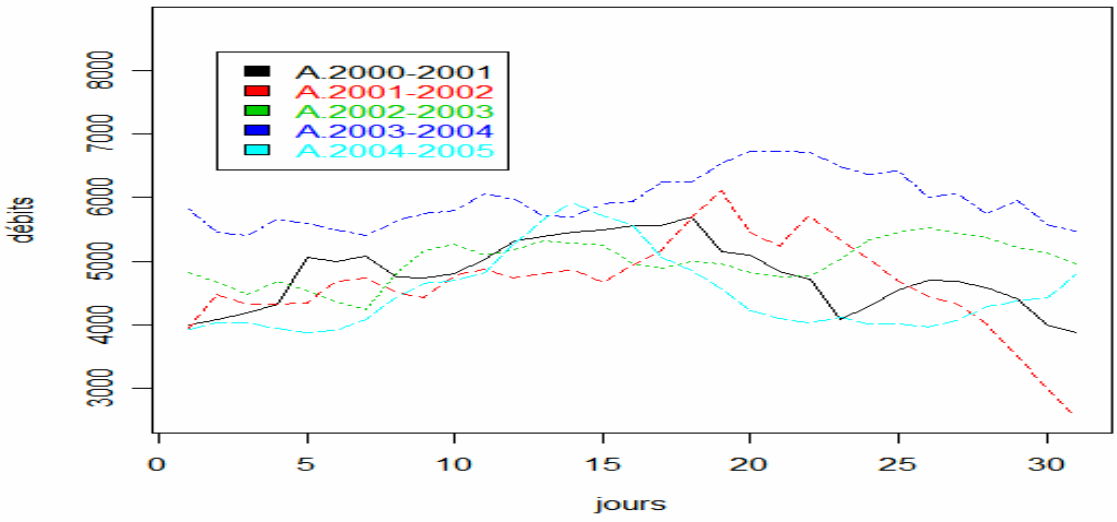


FIG. 2.38 Débits du mois de Novembre par année

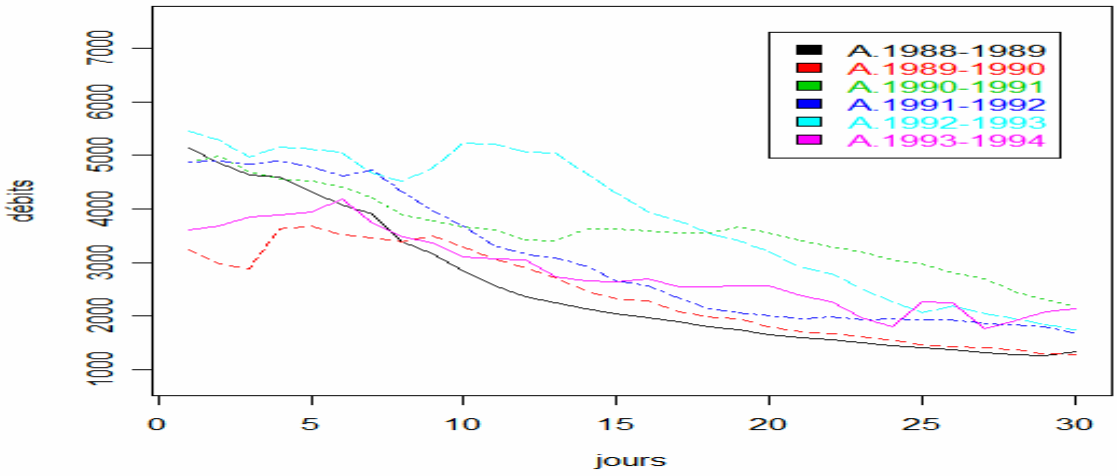


FIG. 2.39 Débits du mois de Novembre par année

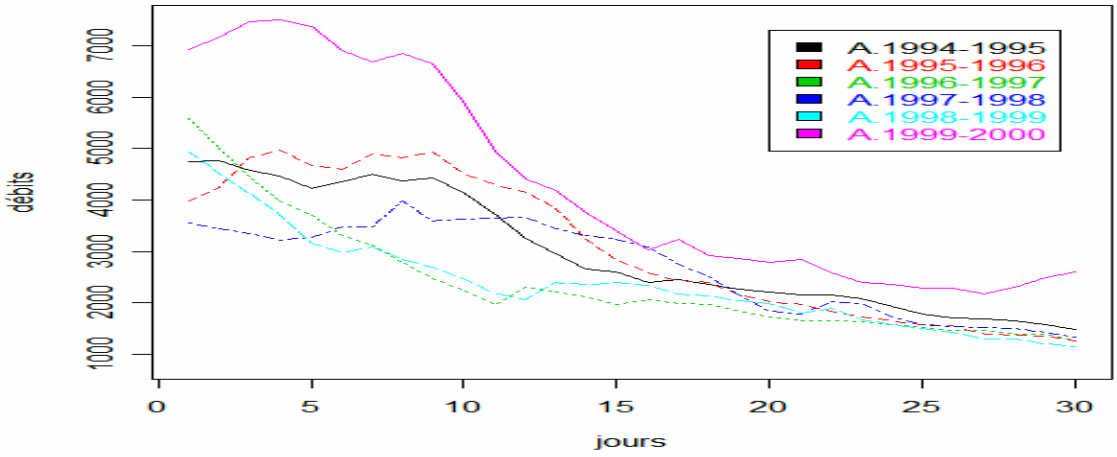
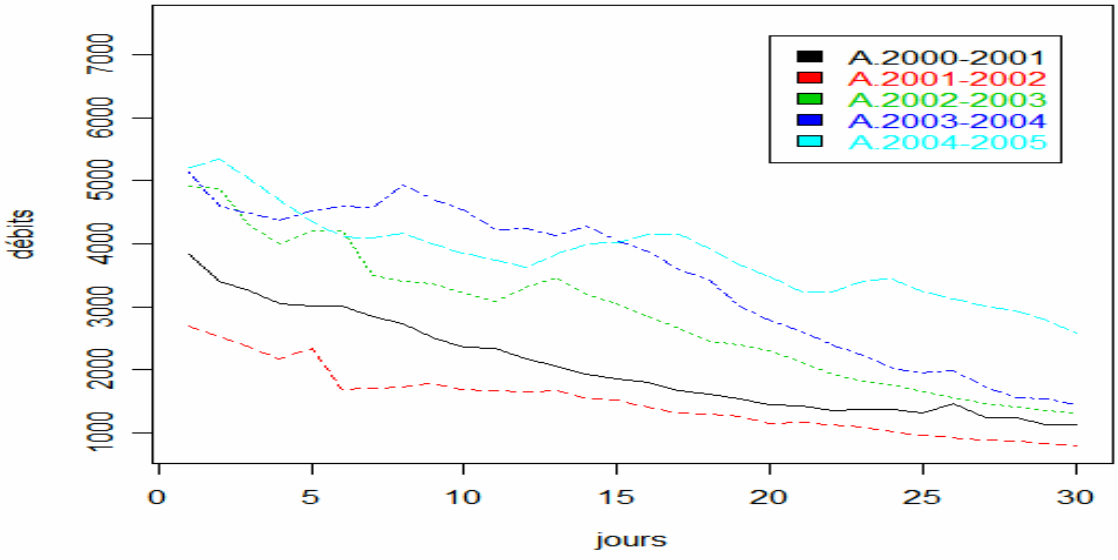


FIG. 2.40 Débits du mois de Novembre par année



Commandes et programmes R utilisés

#==========#
#==========#

- **Lecture des données dans la console du logiciel R**

Les données sont dans un tableau ayant 365 lignes et 17 colonnes. Ce tableau est enregistré dans l'ordinateur, dans le répertoire courant de R, dans un fichier nommé "débits journaliers du bvi.txt". Ce tableau est importé dans la console de R à l'aide de la commande

```
bvi=read.table("débits journaliers du bvi.txt");
```

- C.1 Obtention de la figure 2.1

Nous convertissons le tableau "bvi" en une série temporelle univariée à l'aide de la commande

```
tsbvi=ts(as.vector(as.matirx(bvi)),start=c(1988,1),frequency=365);
```

La fonction suivante permet de tracer la courbe d'évolution d'une série temporelle tout en faisant ressortir éventuellement les différentes périodes à l'intérieur des bandes verticales.

#==========#

```
eda.ts <- function (x, bands=FALSE)  # x est un vecteur ou une série temporelle univariée.
{
       if(!is.ts(x))
      x <- ts(x)
      plot(x,xlab="années",ylab="débits")
      if(bands)
        {
          a <- time(x)
          i1 <- floor(min(a))
          i2 <- ceiling(max(a))
          y1 <- par('usr')[3]
          y2 <- par('usr')[4]
          if( par("ylog") )
            {
              y1 <- 10^y1
              y2 <- 10^y2
            }
         for (i in seq(from=i1, to=i2-1, by=2))
           {
             polygon( c(i,i+1,i+1,i), c(y1,y1,y2,y2), col='grey', border=NA )
           }
         lines(x)
      }
}
```

#==========#

Nous obtenons finalement la figure 2.1 en appliquant successivement les commandes

```
eda.ts(tsbvi,bands=TRUE)
title(main=list("FIG.2.1  Evolution des débits naturels journaliers",col=4,cex=1,font=2))
```

- **C.2 Obtention de la figure 2.2**

Nous utilisons successivement les commandes

```
matbvi=as.matrix(bvi);
matplot(matbvi[,1:6],type="l",xlab="jours",ylab="débits",ylim=range(matbvi));
legend(x=250,y=8000,legend=c("A 1988-1989","A 1989-1990","A 1990-1991","A 1991-1992",
"A 1992-1993","A 1993-1994"),col=1:6,fill=1:6,text.col=1:6);
title(main=list("FIG.2.2  Evolution journalière des débits naturels des années 1988 à 1994",
col=4,cex=0.75,font=2));
```

- **C.3 Obtention de la figure 2.3**

Nous utilisons successivement les commandes

```
 matplot(matbvi[,7:12],type="l",xlab="jours",ylab="débits",ylim=range(matbvi));
legend(x=250,y=8000,legend=c("A 1994-1995","A 1995-1996","A 1996-1997","A 1997-1998",
"A 1998-1999","A 1999-2000"),col=1:6,fill=1:6,text.col=1:6);
title(main=list("FIG.2.3  Evolution journalière des débits naturels des années 1994 à 2000",
col=4,cex=0.75,font=2));
```

- **C.4 Obtention de la figure 2.4**

Nous utilisons successivement les commandes suivantes

```
matplot(matbvi[,13:17],type="l",xlab="jours",ylab="débits",ylim=range(matbvi));
legend(x=250,y=8000,legend=c("A 2000-2001","A 2001-2002","A 2002-2003","A 2003-2004",
"A 2004-2005"),col=1:5,fill=1:5,text.col=1:5);
title(main=list("FIG.2.4  Evolution journalière des débits naturels des années 2000 à 2005",
col=4,cex=0.75,font=2));
```

- **C.5 Obtention de la figure 4.1**

Nous utilisons successivement les commandes suivantes

```
sd=apply(bvi,1,sd)
 moy=apply(bvi,1,mean)
# tableau des débits standardisés #
 bvi.stand=(bvi-moy)/sd
# transformation du tableau des débits standardisés en une série temporelle #
ts.bvi.stand=ts(as.vector(as.matrix(bvi.stand)),start=c(1988,1),frequency=365)
# tracé de la courbe des débits standardisés #
plot(ts.bvi.stand, xlab="années",ylab="débits standardisés")
title(main=list("FIG. 4.1  Evolution des débits standardisés",col=4,cex=1,font=2))
```

- **C.6 Obtention de la figure 4.2**

```
acf(ts.bvi.stand,main="FIG.4.2 Corrélogramme des débits standardisés",lag.max=100)
```

- **C.7 Obtention de la figure 4.3**

```
acf(diff(ts.bvi.stand),main="FIG.4.3 ACF des débits standardisés et différenciés",lag.max=150)
```

- **C.8 Réalisation du test de Phillips-Perron**

```
PP.test(diff(ts.bvi.stand),lshort=F)
```

- **C.9 Obtention de la figure 4.4**

Nous utilisons successivement les commandes suivantes

```
plot(diff(ts.bvi.stand),xlab="années")
title(main=list("FIG. 4.4  Evolution des débits standardisés et différenciés",col=4,cex=1,font=2))
```

- **C.10 Obtention de la figure 4.5**

Nous utilisons successivement les commandes suivantes

```
d=diff(diff(ts.bvi.stand))
acf(d,main="FIG.4.5 ACF des débits standardisés et différenciés 2 fois",lag.max=150)
```

- **C.11 Obtention de la figure 4.6**

pacf(diff(ts.bvi.stand),main="FIG.4.6 PACF des débits standardisés et différenciés",lag.max=150)

- **C.12 Obtention du tableau 4.1**

model1=arima(ts.bvi.stand,order=c(14,1,1))

- **C.13 Obtention du tableau 4.1 (suite)**

round(confint(model1),4)

- **C.14 Obtention du tableau 4.2**

model2=arima(ts.bvi.stand,order=c(13,1,1))

- **C.15 Obtention du tableau 4.2 (suite)**

round(confint(model2),4)

- **C.16 Obtention du tableau 4.3**

model3=arima(ts.bvi.stand,order=c(12,1,1))

- **C.17 Obtention du tableau 4.3 (suite)**

round(confint(model3),4)

- **C.18 Obtention du tableau 4.4**

model4=arima(ts.bvi.stand,order=c(11,1,1))

- **C.19 Obtention du tableau 4.4 (suite)**

round(confint(model4),4)

- **C.20 Obtention de la figure 4.7**

La fonction suivante prend en entrée une série chronologique ou un vecteur ‘z’ et trace le graphe des p-valeurs du test de Ljung-Box en fonction des ordres maximaux de décalages fixés, jusqu’a un certain rang ‘k’, éventuellement précisé.

```
#=============================================================================#
plot.box.ljung <- function (z, k=15, main="p-valeur du test de Ljung-Box")
{
    p <- rep(NA, k)
    for (i in 1:k)
     {
       p[i] <- Box.test(z, i, type = "Ljung-Box")$p.value
     }
   plot(p, type='h', ylim=c(0,1), lwd=3, main=main,xlab="décalages maximaux",
   ylab="probabilités critiques")
   abline(h=c(0,.05),lty=3)
   abline(0.05,0,col=2)
   abline(0.01,0,col=4)
}
#=============================================================================#
```

Nous effectuons successivement les commandes suivantes pour avoir les graphes de la figure 4.7.

```
residus=model4$residuals
par(mfrow=c(2,2))
plot(residus,xlab="années",ylab="résidus",main="Evolution des résidus")
acf(residus,lag.max=150)
pacf(residus,lag.max=150)
plot.box.ljung(residus,k=30)
```

- **C.21 Obtention de la figure 4.8**

Nous effectuons successivement les commandes suivantes

```
residus.stand=(residus-mean(residus))/sd(residus)
hist(residus.stand,proba=TRUE,nclass=100,xlab="résidus standardisés",
main="FIG.4.8 Histogramme des résidus standardisés")
```

- **C.22 Obtention de la figure 4.9**

Nous utilisons successivement les commandes suivantes

```
ts.bvi.stand.est=ts.bvi.stand-residus
ts.bvi.est=(ts.bvi.stand.est*sd)+moy
ts.bvi=ts(as.vector(as.matrix(bvi)),start=c(1988,1),frequency=365)
plot(ts.bvi,xlab="années",ylab="débits")
lines(ts.bvi.est,col=2)
legend(x=1988,y=8500,legend=c("courbe des débits","courbe des débits ajustés"),
text.col=1:2,fill=1:2,col=1:2)
```

- **C.23 Obtention du tableau 4.5**

Nous utilisons successivement les commandes suivantes

```
res=ts.bvi-ts.bvi.est
er=sqrt(sum(res^2)/sum(ts.bvi^2))*100
em=sqrt(sum(res^2)/length(ts.bvi))
```

- **C.24 Obtention du tableau 4.6**

Nous utilisons d'abord le programme suivant

```
#=============================================================================#
 Ce programme prend en entrée par exemple un tableau de débits, prédit sa dernière colonne et
 calcule les erreurs quadratiques de l'ajustement et de la prévision
#=============================================================================#
erreur.prévision.ajustement=function(tab)
{
      tab0=tab[,1:(ncol(tab)-1)]
      # délimitation de la partie du tableau à utiliser pour la ré-estimation#

        sd=apply(tab0,1,sd)
        moy=apply(tab0,1,mean)        #} centrage réduction du tableau des données#
        tab1=(tab0-moy)/sd

        tstab1=ts(as.vector(as.matrix(tab1)),start=c(1988,1),frequency=365)
        # transformation en série temporelle#
        model1=arima(tstab1,order=c(11,1,1))
        # ré-estimation des paramètes du modèle ARIMA(11,1,1)#
```

```
        pred1=predict(model1,n.ahead=365)$pred
        # prévisions de l'année suivante ( dernière colonne du tableau "tab1")#

        ajust1=tstab1-model1$residuals
        ajust1=(ajust1*sd)+moy
        tstab0=ts(as.vector(as.matrix(tab0)),start=c(1988,1),frequency=365)
        res1=tstab0-ajust1 # résidus de l'ajustement#

         pred=(pred1*sd)+moy
         # prévisions de l'année suivante ( dernière colonne du tableau "tab")#
         pred.er=sqrt(sum((tab[,ncol(tab)]-pred)^2)/sum(tab[,ncol(tab)]^2))*100
         pred.em=sqrt(sum((tab[,ncol(tab)]-pred)^2)/365)     #} erreurs quadratiques de la prévision#

         ajust.er=sqrt(sum(res^2)/sum(tstab0^2))*100
         ajust.em=sqrt(sum(res^2)/length(tstab0))   # erreurs quadratiques de l'ajustement#

         sortie=list(prévisions=pred, erreurs=c(er1=ajust.er,,em1=ajust.em,er2=pred.er,em2=pred.em))
         sortie
}
#=========================================================================#
```

puis, nous utilisons successivement les commandes suivantes

```
sortie1=erreur.prévision.ajustement(bvi[,1:14])
sortie2=erreur.prévision.ajustement(bvi[,1:15])
sortie3=erreur.prévision.ajustement(bvi[,1:16])
sortie4=erreur.prévision.ajustement(bvi[,1:17])
```

les différentes lignes de ce tableau sont obtenues en exécutant les instructionsuivantes

```
e1=sortie1[[2]]
e2=sortie2[[2]]
e3=sortie3[[2]]
e4=sortie4[[2]]
```

- **C.25 Obtention de la figure 4.10**

Nous utilisons successivement les commandes suivantes

```
p1=c(sortie1[[1]])
p2=c(sortie2[[1]])
p3=c(sortie3[[1]])
p4=c(sortie4[[1]])
par(mfrow=c(2,2))
plot(bvi[,14],xlab="jours",ylab="débits",main="Année 2001-2002",type="l")
lines(p1,col=4)
legend(x=120,y=6000,legend=c("courbe réelle","courbe prédite"),col=c(1,4),fill=c(1,4),text.col=c(1,4))
plot(bvi[,15],xlab="jours",ylab="débits",main="Année 2002-2003",type="l")
lines(p2,col=4)
legend(x=120,y=6000,legend=c("courbe réelle","courbe prédite"),col=c(1,4),fill=c(1,4),text.col=c(1,4))
plot(bvi[,16],xlab="jours",ylab="débits",main="Année 2003-2004",type="l")
lines(p3,col=4)
legend(x=120,y=6000,legend=c("courbe réelle","courbe prédite"),col=c(1,4),fill=c(1,4),text.col=c(1,4))
plot(bvi[,17],xlab="jours",ylab="débits",main="Année 2004-2005",type="l")
lines(p4,col=4)
legend(x=120,y=6000,legend=c("courbe réelle","courbe prédite"),col=c(1,4),fill=c(1,4),text.col=c(1,4))
```

- **C.26 Obtention de la droite de régression des écarts-types en fonction des moyennes**

Nous utilisons successivement les commandes suivantes

```
ecart=apply(bvi,2,sd)
moyenne=apply(bvi,2,mean)
reg=lm(ecart~moyenne)
r=summary(reg)
```

- **C.27 Obtention de la figure 4.11**

Nous utilisons d'abord le programme suivant

```
#=============================================================================#
Ce programme prend en entrée une série temporelle "serie" et l'ordre de la série moyenne mobile à calculer "p",
et retourne la série des moyennes mobiles centrées.
#=============================================================================#
moy_mob_cent<-function(serie,p)
 {
  if(p%%2!=0)
  {
   cas_imp<-function(serie,p)
    {
     serie_vec<-c(serie)
     n<-length(serie_vec)
     m=p%/%2
     k=n-2*m
     tab=matrix(,nrow=k,ncol=p)
     i=1
     while(p+i-1<=n){tab[i,]=serie_vec[i:(p+i-1)];i=i+1}
     serie_mob1<-apply(tab,1,mean)   # vecteur des moyennes mobiles #
     td<-start(serie)[2]+(p-1)/2
     freq=frequency(serie)
     serie_mob<-ts(serie_mob1,start=c(start(serie)[1]+td%/%freq,td%%freq),frequency=freq)
     # transformation en série temporelle  #
    }
     t<-cas_imp(serie,p)
   }
   else
   {
    cas_pair<-function(serie,p)
     {
     serie_vec<-c(serie)
     n<-length(serie_vec)
     m=p%/%2
     k=n-2*m
     tab=matrix(,nrow=k,ncol=p+1)
     i=1
     while(p+i<=n){tab[i,]=serie_vec[i:(p+i)];i=i+1}
     tab[,c(1,p+1)]=tab[,c(1,p+1)]*0.5
     serie_mob1<-apply(tab,1,function(x){a=sum(x)/p;a})  # vecteur des moyennes mobiles #
     td<-start(serie)[2]+p/2
     freq=frequency(serie)
     serie_mob<-ts(serie_mob1,start=c(start(serie)[1]+td%/%freq,td%%freq),frequency=freq)
     # transformation en série temporelle #
     }
    t<-cas_pair(serie,p)
   }
   t
   }
#=============================================================================#
```

Ensuite, nous utilisons successivement les commandes suivantes

```
serie.moy=moy_mob_cent(ts.bvi,p=365)
plot(ts.bvi,xlab="Années",ylab="débits")
lines(serie.moy,col=2)
legend(x=1988,y=8500,legend=c("courbe des débits","courbe des moyennes mobiles
centrées"),col=1:2,text.col=1:2,fill=1:2)
title(main=list("FIG.4.11 Série des moyennes mobiles d'ordre 365",cex=1,col=4,font=2))
```

- **C.28 Obtention du tableau 4.14**

Nous utilisons d'abord la fonction suivante

```
#=============================================================================#
Ce programme prend en entrée une série temporelle "serie" et retourne les coefficients saisonniers corrigés ainsi
que la série corrigée des variations saisonnières. Ceci en désaisonnalisant cette série à l'aide d'une moyenne
mobile centrée d'ordre "p" et en supposant que le modèle est de type multiplicatif
#=============================================================================#
desaisonnalisation<-function(serie,p)
    {
        rapp_sais_serie<-serie/moy_mob_cent(serie,p)       # calcul des rapports saisonniers #
        c1<-matrix(transform(rapp_sais_serie),ncol=frequency(serie),byrow=TRUE)
        coef_sais<-apply(c1,2,mean,na.rm=TRUE)              # coefficients saisonniers #
        coef_sais_corr<-coef_sais/mean(coef_sais)           # coefficients saisonniers corrigés #
        cvs<-serie/coef_sais_corr                           # série corrigée des variations saisonnières #
         sortie=
      list(Coefficients_Saisonniers_Corrigés=coef_sais_corr,Série_Corrigée_Des_Variations_Saisonnières=cvs)
        sortie
    }
#=============================================================================#
```

Puis, la commande suivante

```
des=desaisonnalisation(ts.bvi,p=365)
des[[1]]
```

- **C.29 Obtention de la figure 4.12**

Nous avons effectués successivement les commandes suivantes

```
plot(ts.bvi,xlab="années",ylab="débits")
lines(des[[2]],col=2)
t=1:length(des[[2]]
t=1:length(des[[2]])
reg1=lm(c(des[[2]])~t)
abline(reg1,col=4)
legend(x=1988,y=8500,legend=c("courbe des débits","courbe de la série corrigée des variations
saisonnières","courbe de la tendance globale estimée"),text.col=c(1,2,4),col=c(1,2,4),fill=c(1,2,4))
```

- **C.30 Obtention de la tendance** C_t

```
s=summary(reg1)
```

- **C.31 Obtention de la figure 4.13**

Nous utilisons successivement les commandes suivantes

```
res.est=c(ts.bvi)/rep(reg1$coefficients[1]*des[[1]],17)
ts.res.est=ts(as.vector(as.matrix(res.est)),start=c(1988,1),frequency=365)
plot(ts.res.est,xlab="années",ylab="résidus",main="FIG.4.13 Graphe des résidus")
```

- **C.32 Obtention de la figure 4.14**

```
acf(ts.res.est,lag.max=150,main="FIG.4.14  Résidus estimés")
```

- **C.33 Obtention de la figure 4.15**

```
acf(diff(ts.res.est),lag.max=150,main="Résidus différenciés")
pacf(diff(ts.res.est),lag.max=150,main="Résidus différenciés")
```

- **C.34 Obtention du tableau 4.7**

```
model3=arima(tsr2,order=c(16,1,1))
```

- **C.35 Obtention du tableau 4.7 (suite)**

```
round(confint(model3),4)
```

- **C.36 Obtention de la figure 4.16**

Nous utilisons les commande suivantes

```
plot(model3$residuals,xlab="Années",ylab="résidus",main="Courbe des résidus")
acf(model3$res,main="Résidus",lag.max=150)
pacf(model3$res,main="Résidus",lag.max=150)
plot.box.ljung(model3$res,k=30)
```

- **C.37 Obtention de la figure 4.17**

Nous utilisons successivement les commandes suivantes

```
res.est1=ts.res.est-model3$res
ts.bvi.est1=reg1$coef[1]*rep(des[[1]],17)*res.est1
plot(ts.bvi,xlab="Années",ylab="débits")
lines(ts.bvi.est1,col=2)
legend(x=1988,y=8500,legend=c("courbe des débits réels","courbes des débits
estimés"),col=1:2,text.col=1:2,fill=1:2)
```

- **C.38 Obtention du tableau 4.8**

Nous utilisons les commandes suivantes

```
r=ts.bvi-ts.bvi.est1
er=sqrt(sum(r^2)/sum(ts.bvi^2))*100
em=sqrt(sum(r^2)/6205)
```

- **C.39 Obtention du tableau 4.9**

Nous utilisons d'abord le programme suivant

```
#=============================================================================#
Ce programme prend en entrée par exemple un tableau de débits, prédit sa dernière colonne et
 calcule les erreurs quadratiques de l'ajustement et de la prévision. Ceci en modélisant la chronique obtenue
avec le reste des données de ce tableau par un modèle multiplicatif, tout en désaisonnalisant cette chronique
par la méthode des moyennes mobiles centrées.
#=============================================================================#
erreur.prévision.ajustement1=function(tab)
{
  k=ncol(tab)-1
  tab0=tab[,1:k]
 # délimitation de la partie du tableau à utiliser pour la ré-estimation#
```

```
tstab0=ts(as.vector(as.matrix(tab0)),start=c(1988,1),frequency=365)
# transformation en série temporelle#

des=desaisonnalisation(tstab0,p=365)
#ré estimation des coefficients saisonniers et de la série corrigéé des variations saisonnières #
cvs=c(des[[2]])
t=1:length(cvs)
reg=lm(cvs~t)
tstab0.ajust1=rep(reg$coef[1]*des[[1]],k)
tstab0.ajust1=ts(as.vector(as.matrix(tstab0.ajust1)),start=c(1988,1),frequency=365)
# calcul de la série estimée par le modèle multiplicatif #

ts.res1=tstab0/tstab0.ajust1
model=arima(ts.res1,order=c(16,1,1))
#ré-estimation des paramètres du modèle ARIMA(16,1,1) #
pred1=predict(model,n.ahead=365)$pred
# prévision de l'année suivante par le modèle ARIMA(16,1,1) #
pred=reg$coef[1]*des[[1]]*pred1
# prévision de l'année suivante (dernière colonne du tableau "tab") #

ts.res.ajust=ts.res1-model$res
tstab0.ajust=tstab0.ajust1*ts.res.ajust  # ajustement finale de la chronique utilisée pour la ré-estimation #
res=tstab0-tstab0.ajust  # résidus de l'ajustement finale #

pred.er=sqrt(sum((tab[,ncol(tab)]-pred)^2)/sum(tab[,ncol(tab)]^2))*100
pred.em=sqrt(sum((tab[,ncol(tab)]-pred)^2)/365)     #} erreurs quadratiques de la prévision#

ajust.er=sqrt(sum(res^2)/sum(tstab0^2))*100
ajust.em=sqrt(sum(res^2)/length(tstab0))   # erreurs quadratiques de l'ajustement#

sortie=list(prévisions=pred, erreurs=c(er1=ajust.er,,em1=ajust.em,er2=pred.er,em2=pred.em))
sortie
}
#=============================================================================#
```

Puis successivement les commandes suivantes

```
sortie1=erreur.prévision.ajustement1(bvi[,1:14])
sortie2=erreur.prévision.ajustement1(bvi[,1:15])
sortie3=erreur.prévision.ajustement1(bvi[,1:16])
sortie4=erreur.prévision.ajustement1(bvi[,1:17])
e1=sortie1[[2]]
e2=sortie2[[2]]
e3=sortie3[[2]]
e4=sortie4[[2]]
```

- **C.40 Obtention de la figure 4.18**

Nous utilisons successivement les commandes suivantes

```
p1=c(sortie1[[1]])
p2=c(sortie2[[1]])
p3=c(sortie3[[1]])
p4=c(sortie4[[1]])
par(mfrow=c(2,2))
plot(bvi[,14],xlab="jours",ylab="débits",main="Année 2001-2002",type="l")
lines(p1,col=4)
legend(x=120,y=6000,legend=c("courbe réelle","courbe prédite"),col=c(1,4),fill=c(1,4),text.col=c(1,4))
plot(bvi[,15],xlab="jours",ylab="débits",main="Année 2002-2003",type="l")
lines(p2,col=4)
legend(x=120,y=6000,legend=c("courbe réelle","courbe prédite"),col=c(1,4),fill=c(1,4),text.col=c(1,4))
```

```
plot(bvi[,16],xlab="jours",ylab="débits",main="Année 2003-2004",type="l")
lines(p3,col=4)
legend(x=120,y=6000,legend=c("courbe réelle","courbe prédite"),col=c(1,4),fill=c(1,4),text.col=c(1,4))
plot(bvi[,17],xlab="jours",ylab="débits",main="Année 2004-2005",type="l")
lines(p4,col=4)
legend(x=120,y=6000,legend=c("courbe réelle","courbe prédite"),col=c(1,4),fill=c(1,4),text.col=c(1,4))

#=========================================================================#
```

- **C.41 Obtention de la figure 4.19**

Nous supposerons ici qu'on a un vecteur nommé « debitmoy » contenant la série des débits moyens mensuels commençants à partir du moi de Juillet.
Nous utilisons les commandes suivantes

```
ts.debit.moy=ts(as.vector(as.matrix(t(debitmoy))),start=c(1988,1),frequency=12)
eda.ts(ts.debit.moy,bands=TRUE)
title(main=list("FIG. 4.19  Courbe des débits moyens mensuels",cex=1,col=4,font=2))
```

- **C.42 Obtention de la figure 4.20**

Nous utilisons les commandes suivantes

```
par(mfrow=c(2,1))
acf(ts.debit.moy,lag=50,main="Auto-corrélogramme")
pacf(ts.debit.moy,lag=50,main="Auto-corrélogramme partiel")
```

- **C.43 Obtention de la figure 4.21**

Nous utilisons les commandes suivantes

```
Par(mfrow=c(2,1))
acf(diff(ts.debit.moy,main="Auto-corrélogramme"lag=50)
pacf(diff(ts.debit.moy,main= "Auto-corrélogramme partiel",lag=50)
```

- **C.44 Test de Phillips-Perron sur le processus Z_t**

```
PP.test(diff(ts.debit.moy,lag=12))
```

- **C.45 Obtention du tableau 4.10**

```
model4=arima(ts.debit.moy,order=c(1,0,2),seasonal=list(order=c(1,1,1),period=12))
```

- **C.46 Obtention du tableau 4.10 (suite)**

```
round(confint(model4),4)
```

- **C.47 Obtention du tableau 4.11**

```
model5=arima(ts.debit.moy,order=c(1,0,1),seasonal=list(order=c(1,1,1),period=12))
```

- **C.48 Obtention du tableau 4.11 (suite)**

```
round(confint(model5),4)
```

- **C.49 Obtention du tableau 4.12**

```
model6=arima(ts.debit.moy,order=c(1,0,0),seasonal=list(order=c(1,1,1),period=12))
```

- **C.50 Obtention du tableau 4.12 (suite)**

```
round(confint(model6),4)
```

- **C.51 Obtention de la figure 4.22**

Nous utilisons successivement les commandes suivantes

```
par(mfrow=c(2,2))
plot(model6$res,xlab="mois",ylab="débits",main="Courbe des résidus")
acf(model6$res,main="Résidus",lag=36)
pacf(model6$res,main="Résidus",lag=36)
plot.box.ljung(model6$res,k=30)
```

- **C.52 Obtention de la figure 4.23**

```
plot(ts.debit.moy,xlab="Années",ylab="débits")
lines(ts.debit.moy-model6$res,col=2)
legend(x=1988,y=7000,legend=c("courbe des débits moyens mensuels","courbe des débits moyens
ajustés"),col=1:2,text.col=1:2,fill=1:2)
title(main=list("FIG. 4.23 Ajustement des débits moyens mensuels",cex=1,col=4,font=2))
```

- **C.53 Obtention du tableau 4.13**

Nous utilisons successivement les commandes suivantes

```
er=sqrt(sum(model6$res^2)/sum(ts.debit.moy^2))*100
em=sqrt(sum(model6$res^2)/length(ts.debit.moy))
```

- **C.54 Programme de prévision**

```
#=============================================================================#
 Ce programme prend en entrée par exemple un tableau de débits, prédit les débits de l'anné suivante.
#=============================================================================#
programme.prédiction=function(tab)

{

 sd=apply(tab,1,sd)
 moy=apply(tab,1,mean)        #} centrage réduction du tableau des données#
 tab1=(tab-moy)/sd

tstab1=ts(as.vector(as.matrix(tab1)),start=c(1988,1),frequency=365)  # transformation en série temporelle#
model1=arima(tstab1,order=c(11,1,1))             # estimation des paramètes du modèle ARIMA(11,1,1)#
pred1=predict(model1,n.ahead=365)$pred      # prévisions de l'année suivante par le modèle
ARIMA(11,1,1)#

pred=(pred1*sd)+moy       # prévisions de l'année suivante ( dernière colonne du tableau "tab")#
pred=as.data.frame(c(pred))
pred

}

#=============================================================================#
```

#==#

- **C.55 Programme de simulation**

#==#

Ce programme R permet de simuler les débits naturels journaliers du B.V.I. de la Sanaga sur une période donnée. Il prend en entrée 4 paramètres:

1- tab est un tableau de 365 lignes qui représentent les jours d'une année (la première ligne contient les débits du 1er Juillet et dont les colonnes représentent les années réellement observées qui serviront à l'estimation des coefficients du modèle;
2- qf est la valeur qui précède la première valeur à simuler;
3- tf est le numéro du jour de l'année où l'on désire commencer la simulation;
4- n est le nombre de jours des débits à simuler.

#==

```
programme.simulation=function(tab,qf,tf,n)
{
  QH=apply(tab,1,mean)
  QH=c(QH[365],QH)              #} estimation des moyennes des débits journaliers

  SG=apply(tab,1,sd)
  SG=c(SG[365],SG)              #} estimation des écarts-types des débits journaliers

  EPS=rnorm(n)                  # génération de l'impulsion aléatoire

  AC=c()
  QS=c()

  for(i in 1:364)
   {
    AC[i]=cor(t(tab[i,]),t(tab[i+1,]))    # calcul des 364 premiers coefficients d'auto-corrélation
   }

   AC=c(AC,cor(t(tab[365,]),t(tab[1,])))   # la dernière valeur de ce vecteur est le coefficient
                                           # d'autocorrélation entre entre le 1er et le dernier  jour de l'année
   AC=c(AC[364],AC[365],AC[1:364])

   QS[1]=qf
   j=tf
   i=1

   while(i<n+1)
    {
     QS[i+1]=QH[(j%%365)+1]+(AC[(j%%365)+1]*SG[(j%%365)+1]/SG[(j-1)%%365+1])*(QS[i]-QH[(j-
1)%%365+1])+EPS[i]*SG[(j%%365)+1]*sqrt(1-AC[(j%%365)+1]^2)
      i=i+1
      j=j+1
    }

   QS=QS[-1]

   plot(QS,type="l",xlab="jours",ylab="débits")
   title(main=list(" Courbe des débits journaliers simulés",cex=1,font=2,col=4))
   QS=as.data.frame(QS)
   QS
  }
```

#==#

- **C.56 Obtention des figures 4.24 et 4.25**

Nous effectuons successivement les commandes suivantes :

```
s=programme.simulation(tab=bvi,qf=1400,tf=1,n=
s=programme.simulation(tab=bvi,qf=1400,tf=1,n=6205)
```

TAB. 4.14 : Coefficients saisonniers corrigés

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre
1	0.3796690	0.18493876	0.08587659	0.1057755	0.2697603	0.4951706	0.7530991	1.317714	2.002241	2.687462	2.6032219	0.8732004
2	0.3761328	0.17933059	0.08419594	0.1112016	0.2802059	0.5105538	0.7607697	1.345310	1.974055	2.695277	2.5368281	0.8521936
3	0.3635125	0.17512521	0.08767996	0.1118542	0.2733742	0.5113827	0.7963744	1.362327	2.043516	2.709429	2.4591116	0.8128188
4	0.3508398	0.17490494	0.08692630	0.1143066	0.2931471	0.5444097	0.8299763	1.409952	2.087100	2.730005	2.4181327	0.7856386
5	0.3463216	0.16469434	0.08386257	0.1135146	0.2999013	0.5291337	0.8511784	1.426061	2.129695	2.763001	2.3685970	0.7569524
6	0.3315193	0.17137307	0.08297303	0.1185111	0.3082965	0.5402105	0.8790408	1.436476	2.161386	2.802147	2.2996309	0.7282198
7	0.3252176	0.15495350	0.08424114	0.1217768	0.3100021	0.5526233	0.8941734	1.471021	2.220847	2.834839	2.2445962	0.7042345
8	0.3222885	0.15210029	0.08587963	0.1339892	0.3355491	0.5489923	0.9129189	1.522616	2.278948	2.841860	2.2115854	0.6836169
9	0.3137040	0.14082236	0.08443435	0.1319943	0.3328437	0.5679330	0.9049363	1.552810	2.319541	2.846886	2.1497290	0.6652021
10	0.3062055	0.13622946	0.07882675	0.1462823	0.3553160	0.5626541	0.9149925	1.598358	2.369724	2.845882	2.0460343	0.6300033
11	0.2963720	0.12789888	0.07550963	0.1519665	0.3497938	0.5733967	0.9174243	1.640178	2.450489	2.851674	1.9291075	0.6149559
12	0.2796069	0.12700800	0.07546442	0.1618654	0.3545699	0.5830151	0.9447402	1.661589	2.532675	2.874701	1.8729537	0.5964258
13	0.2729070	0.13246670	0.07326114	0.1586335	0.3492717	0.5809153	0.9607911	1.730141	2.490435	2.911207	1.8261835	0.5863441
14	0.2710539	0.12691137	0.07963552	0.1749035	0.3553345	0.5794510	0.9612762	1.738214	2.505495	2.987601	1.7359054	0.5840215
15	0.2644890	0.11624821	0.08555810	0.1792428	0.3705559	0.5937129	0.9583124	1.695944	2.565288	2.983950	1.6548473	0.5665892
16	0.2636139	0.11338837	0.09396589	0.1878837	0.3807479	0.5991211	0.9590220	1.721787	2.579796	2.949320	1.5876058	0.5603451
17	0.2577130	0.10725076	0.09989598	0.1838451	0.3758604	0.6153249	0.9816387	1.776321	2.620607	2.966830	1.5141602	0.5625515
18	0.2513348	0.10881096	0.09523547	0.1902555	0.3775258	0.6082985	0.9955796	1.830646	2.648335	3.041485	1.4460294	0.5292506
19	0.2340351	0.10988175	0.10241467	0.1855213	0.3958276	0.6146208	1.0052062	1.827777	2.608349	3.043789	1.3741714	0.5166074
20	0.2357549	0.10232998	0.09951134	0.2030447	0.4145044	0.6292879	1.0229634	1.845726	2.597904	3.037069	1.3009844	0.5120561
21	0.2282904	0.10013647	0.10430767	0.2055439	0.4308500	0.6414542	1.0580221	1.829238	2.608833	3.031778	1.2392199	0.4871893
22	0.2297083	0.09950772	0.11244868	0.2074276	0.4375365	0.6402054	1.0694532	1.848485	2.599073	3.019597	1.2088604	0.4767820
23	0.2180623	0.09732462	0.11516615	0.2190924	0.4342693	0.6572354	1.0991195	1.880121	2.572153	2.931348	1.1577355	0.4705400
24	0.2167725	0.09738838	0.12486925	0.2326665	0.4445290	0.6661133	1.1370143	1.912336	2.614463	2.911541	1.1001274	0.4556433
25	0.2110205	0.09259230	0.11744997	0.2307602	0.4530725	0.6728180	1.1328172	1.921067	2.643617	2.918597	1.0610290	0.4475260
26	0.2090153	0.09459257	0.11279848	0.2313367	0.4763008	0.6887207	1.1459181	1.947258	2.626748	2.902766	1.0478197	0.4392489
27	0.2041904	0.08778002	0.10864843	0.2373260	0.4890303	0.7101340	1.1581093	2.012211	2.669540	2.790362	0.9769205	0.4291174
28	0.2023214	0.08815043	0.11222525	0.2583998	0.4942380	0.7277065	1.2003363	2.023916	2.669323	2.750257	0.9546143	0.4179543
29	0.1981582		0.11132725	0.2587246	0.4851266	0.7403042	1.2774254	2.027334	2.624087	2.734349	0.9307758	0.4125014
30	0.1931589		0.10621591	0.2621816	0.4859664	0.7453815	1.2718017	1.997396	2.641230	2.672499	0.8922683	0.4108620
31	0.1854052		0.11885289		0.4783964		1.2919044	2.003036		2.681767		0.3941446

Bibliographie

[1] ARTHUR Charpentier (2004). '*Série Temporelles - théorie et applications'* polycopié de cours.

[2] BOX & JENKINS (2002). '*Application des modèles de type « ARIMA » en hydrologie'*

[3] COYNE & BELLIER (2002). '*Arbitrage Alucam / AES-SONEL : Rapport de mission'*

[4] DROESBEKE, J. J, FICHET, B & TASSI, P. (1995). *'Séries chronologiques – théorie et pratique des modèles ARIMA',* Economica

[5] MAYAP Marie Adèle (2005). '*Analyse Statistique et Calcul des Probabilités d'Incidents : cas du Réseau Interconnecté Sud de AES-SONEL*.' Mémoire de master de statistique appliquée, ENSP, Yaoundé.

[6] M.DANECH-PA J OUH (2003). *'Les modèles de prévision du dispositif de Bison futé et leur évolution'*. INRETS : Institut National de Recherche sur le Transport et leur Sécurité.

[7] NINO Sylverio (2005). *'Séries chronologiques'*. Polycopié de cours.

[8] Organisation Météorologique Mondiale (5ème édition 1994). *Guide des pratiques Hydrologiques*

[9] P. DUBREUIL, J. C. OLIVRY, J. GUISCAFRE, J. F. NOUVELOT (1975). '*Monographie Hydrologique ORSTOM : le bassin de la rivière Sanaga*', Paris Vol 3**.**